

Visual Thought

Advances in Consciousness Research

Advances in Consciousness Research provides a forum for scholars from different scientific disciplines and fields of knowledge who study consciousness in its multifaceted aspects. Thus the Series will include (but not be limited to) the various areas of cognitive science, including cognitive psychology, linguistics, brain science and philosophy. The orientation of the Series is toward developing new interdisciplinary and integrative approaches for the investigation, description and theory of consciousness, as well as the practical consequences of this research for the individual and society.

Editor

Maxim I. Stamenov
Bulgarian Academy of Sciences

Editorial Board

David Chalmers
Australian National University

Gordon G. Globus
University of California at Irvine

Ray Jackendoff
Brandeis University

Christof Koch
California Institute of Technology

Stephen Kosslyn
Harvard University

Earl Mac Cormac
Duke University

Steven Macknik
Barrow Neurological Institute

George Mandler
University of California at San Diego

Susana Martinez-Conde
Barrow Neurological Institute

John R. Searle
University of California at Berkeley

Petra Stoerig
Universität Düsseldorf

Volume 67

Visual Thought: The depictive space of perception
by Liliana Albertazzi

Visual Thought

The depictive space of perception

Liliana Albertazzi

Trento University

·
·

John Benjamins Publishing Company

Amsterdam/Philadelphia



™ The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Visual thought : the depictive space of perception / [edited by] Liliana
Albertazzi.

p. cm. (Advances in Consciousness Research, ISSN 1381-589X ; v.
67)

Includes bibliographical references and indexes.

1. Space perception. 2. Consciousness. I. Albertazzi, Liliana.

BF469.V57 2006

152.14--dc22

2006049946

ISBN 90 272 5203 3 (Hb; alk. paper)

© 2006 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

To John Willats
In memoriam

Table of contents

Affiliations, addresses	IX
Foreword	XI

Part I. Perception of visual spaces

Introduction to visual spaces <i>Liliana Albertazzi</i>	3
Coplanar reflectance change and the ontology of surface perception <i>Dhanraj Vishwanath</i>	35
Evidence suggestive of separate visual dynamics in perception and in memory <i>Timothy L. Hubbard and Jon R. Courtney</i>	71
Illusory space and paradoxical transparency in stereokinetic objects <i>Mario Zanforlin</i>	99
The neural space of visual shape perception <i>Ed Connor</i>	105
Boundary Gestalt limits flow Gestalt: The geometry of good continuation <i>Ohad Ben-Shahar and Steven W. Zucker</i>	115

Part II. Depiction of visual spaces

Pictorial space, a modern reappraisal of Adolf Hildebrand <i>Jan J. Koenderink and Andrea J. van Doorn</i>	135
Gestalts of thought <i>Barbara Tversky</i>	155
Visual quality: Drawing on canvas <i>Liliana Albertazzi</i>	165

Rudolf Arnheim's graphic equivalents in children's drawings and drawings and paintings by Paul Klee <i>John Willats</i>	195
Visual perception and theories of painting: An uneasy complementarity <i>Alf C. Zimmer</i>	221
 Part III. Bridging perception and depiction of visual spaces	
Dynamics of picture viewing and picture description <i>Jana Holsanova</i>	235
Order and complexity in naturalistic landscapes: On creation, depiction and perception of Japanese dry rock gardens <i>Gert J. van Tonder</i>	257
Thoughts on shape <i>Frederic Fol Leymarie</i>	303
Tracing axes of growth <i>Athanassios Economou</i>	351
John Willats†	367
Name index	373
Subject index	377

Affiliations, addresses

Liliana Albertazzi
Department of Cognitive Science
and Education
Matteo del Ben Street, 5
Rovereto, Italy
liliana.albertazzi@unitn.it

Ohad Ben-Shahar
Department of Computer Science
Ben Gurion University
Beer-Sheva, Israel
ben-shahar@es.tgu.ac.il

Ed Connor
John Hopkins University
3400 N. Charles St.
Baltimore, MD 21218, USA
connor@jhu.edu

Athanassios Economou
College of Architecture
Georgia Institute of Technology
Atlanta, Georgia 30305, USA
economou@coa.gatech.edu

Jana Holsanova
Cognitive Science Department
Lund University
Kungshuset, Lundagård
222 22 Lund, Sweden
jana.holsanova@lucs.lu.se

Thimothy L. Hubbard
Department of Psychology
Texas Christian University
Fort Worth, TX 76129, USA
t.hubbard@tcu.edu

Jan J. Koenderink D.Sc.
Physics & Astronomy
Buys Ballot Laboratory
Princetonplein 5
3584CC Utrecht, The Netherlands
j.j.koenderink@phys.uu.nl

Frederic Fol Leymarie
Computing Department
Goldsmiths College University
of London, UK
ff@gold.ac.uk

Barbara Tversky
Columbia Teachers College
525 W. 20th St.
New York, NY 10027, USA
bt@psych.stanford.edu

Andrea J. van Doorn
Department of Physics of Man
Helmholtz Institute, Utrecht University
Princeton Plein 5
3584 CC Utrecht, The Netherlands
a.koenderink@phys.uu.nl

Gert van Tonder
Department of Architecture & Design
East bldg 1 Room 509 Kyoto
Institute of Technology
Matsugasaki, Kyoto 606-8585, Japan
gvtonder@yahoo.co.uk

Dhanraj Vishwanath
Department of Psychology
Rochester Institute of Technology
18 Lomb Memorial Drive
Rochester, NY 14607, USA
dxvgss@rit.edu

Mario Zanforlin
Department of General Psychology
Padova University
Venezia Street, 8
35100 Padova, Italy
mario.zanforlin@unipd.it

Alfred Zimmer
Department of Psychology

University Regensburg
93040 Regensburg, Germany
alf.zimmer@psychologie.uni-regensburg.de

Steve W. Zucker
Department of Computer Science
Yale University
New Have, CT 06520, USA
steven.zucker@yale.edu

Foreword

This volume is a first result of an ongoing research project on form analysis which began some years ago at Mitteleuropa Foundation in Bolzano, Italy (www.mittleeuropafoundation.org) and has provided the framework for stimulating work.

The research group on form analysis is an interdisciplinary research group in Cognitive Science which works on visual perception. The group systematically acknowledges in its program the legacy of scientific phenomenology and Gestalt psychology.

The scholars currently involved in the project – Athanassios Economou (Georgia Institute of Technology, Atlanta), Ernest Edmonds (Sydney University), Fred-eric Fol Leymarie (Goldsmiths College, London), Michael Leyton (Rutgers University and D.I.M.A.C.S.), Nancy Nersessian (Georgia Institute of Technology), Gert van Tonder (Kyoto Institute of Technology), John Willats (Loughborough University), Dhanraj Vishwanath (Rochester Institute of Technology), and myself (Trento University and Mitteleuropa Foundation) – have met on a regular basis since 2003. Other scholars are closely connected with the group, and some of them are contributors to the present volume. One of the main purposes of the project is to analyse the relation and similarity between perceptual space and the depictive space of art.

More generally, the group is working on a new research paradigm for perceptual science which starts from how we ‘experience’ vision and seeks to identify the best tools with which to model that type of complexity. In other words, the group is working on the very *origins of conscious qualitative states* as a major aspect of information theory. However ambitious, the project (experience to neurons and stimuli!) is producing stimulating ideas and innovative research, and it is progressively gaining consensus among scholars in different disciplinary fields.

Specifically, this volume originates from a 2004 event on form perception and understanding (*The Depictive Space of the Mind*, see Mitteleuropa Foundation web site under ‘Past events’), devoted to the primary level of visual appearances in the phenomenal field. As its main goal, the volume examines the complexity of the field of vision from an original viewpoint, conjugating the principles of Gestalt psychology with more recent scientific insights. Tracing back the origins of a conceptual legacy, both theoretical and experimental, embedded in Gestalt psy-

chology, in fact, is not an exercise in nostalgia but intends to contribute to the real advancement of contemporary research.

The volume is organized as following. The first chapter (Albertazzi) presents the general framework of the research topic, i.e. the co-presence of diverse types of spaces in vision, as the optical space of psychophysics and of neural elaboration, the qualitative space of phenomenal appearances, and the pictorial space of art. It argues that those spaces follow different rules of organization, whose specific singularity and reciprocal dependence must be individuated as a preliminary step, before conceiving their implementation in cognitive agents. Other contributions are devoted to the following: specific aspects of the topic concerning primary and secondary cognitive processes, analyzed on the basis of the Gestalt principles from both a conceptual and an experimental point of view (Viswhanath, Tversky and Hubbard); the dynamics of neural space, and in particular the role of the law of good continuations in forming the boundary of phenomenal forms (Zucker, and Connor); the relation between perceptual and pictorial space, in theory and experimentation (Zimmer, Albertazzi, Koenderink/van Doorn, Willats and Zanforlin); extending the paradigm to landscape configuration and perception (van Tonder), and to the relation between pictorial perception and linguistic description (Holsanova); finally, reflection on form modelling from the point of view of the different geometries involved (Leymarie), and from the point of view of design (Economou).

I thank all the contributors to the volume for being part of an exciting enterprise in scientific and aesthetic progress, and Mitteleuropa Foundation for providing such a stimulating environment for our work.

Liliana Albertazzi
Trento University and Mitteleuropa Foundation
Trento/Bolzano 2006

PART I

Perception of visual spaces

Introduction to visual spaces

Liliana Albertazzi

Spaces' dynamics

A leading issue in contemporary theory of vision is the difference among *optical* space, the *phenomenal* space of appearances, and *pictorial* space, which in principle concerns different kinds of 'spaces' but displays several connections from the conceptual, the experimental and the methodological point of view (to mention only some (very different) approaches to the problem of vision see: Gibson 1979; Rock 1983; Knill, Richards 1996; Mausfeld 2002).

However, these general distinctions are not always straightforward. For example the structure of optical space has usually been considered a structure of metric type and has consequently been identified with physical space. But recent studies tend to treat optical space as a sort of *homogeneous* space based not on metric relations but on relations of incidence. This change of view has had important consequences: for example, optical space loses its characteristics of veridicality (Koenderink, van Doorn, Kappers, Todd 2002. On veridicality in perception see Palmer 1999:6).

Matters become even more complex when one considers that each kind of perceptual *act* (seeing, hearing, touching, smelling) has inherent it, or embedded in it, a specific 'space' which cooperates in the intermodality of the senses: for example, there is experimental evidence for visual and proprioceptive contributions to tactile localization (Hornbostel 1925; Koffka 1930; Cytowic 1989; Rich, Mattingley 2002). There are then different kinds of modules of these spaces: for example, in the space of vision scientists distinguish among 'personal space', 'action space' and 'visual space' in the proper sense, which indicates that there is an integrated system at work (Cutting, Vishton 1995; Làvadas, Pavani 1998; Làvadas 2002; Botvinick, Cohen 1998; Albertazzi 2002b; Spence, Driver 2004; Holmes, Calvert, Spence forthcoming; Holmes, Spence 2004).

Also to be considered is the fact that vision is the sensory modality with the greatest spatial acuteness, and as such it is able to modulate other perceptive modalities: for example, many aspects of auditory perception are influenced by

the localization of a visual stimulus (Lewald, Ehrenstein 1996). In the specific case of peripersonal visual space, the space relative to a limb (a hand, for example) is contracted or expanded according to the tool use, in other situations, the entire peripersonal space of the body may be deceived by the vision of an artificial limb (Holmes et al., submitted). From this one may deduce that, in the first instance, visual space, and more generally perceived space as a whole, are embedded in a *network structure*. This would also suggest that one should not expect only a single metric in visual space, be it Euclidean or not (Todd, Oomes, Koenderink, Kappers 2001 suggest that visual space satisfies the weaker constraints of affine geometrical structure).

These preliminary considerations immediately raise the question of the relation between these *spaces* and their *objects*. The issue in particular is whether space should be understood as a container or instead, in certain cases, as a constitutive, and therefore non-independent, part of objects in interaction. In the phenomenal visual field, for example, objects and/or groups of objects cannot be treated *à la* Zeno as atomic individuals in a space/container because, as the Gestalt analysis has amply shown, they are *effects* of the organization of perception (Koffka 1935, Chapter 3; Kopfermann 1930).¹

But one need only consider such elementary concepts as *extension*, *distance* or *direction* in the visual field to be aware of the coexistence and interdependence of multiple 'spaces' and of the numerous factors involved in the construction of the objects of vision.

In the case of the perception of *distance*, for instance, converging ocular movements, variations of form in linear perspective, occlusion, qualitative differences in shape, colour and brightness, textures of surfaces, line orientation, movement of the observer or of part of the scene, height in the visual field, distribution of attention, relative magnitude and variation in the magnitude of the objects known, are all factors which position an object as 'close' or 'distant' in visual space. Even the preference for geometric forms, planar surfaces or repetition of patterns in natural and artificial environment matters in perceiving distance.² Moreover, it is well known that the principle of distance is one of the factors that determine the figural grouping and is therefore a 'completion' of the objects in the visual field (Wertheimer 1923); and also that variations in distance are more visible in a small space than in a large one, with the consequence that variations in distance influence the perception of movement (Duncker 1929; Koffka 1935:290).

In the case of moving objects, moreover, the perception of distance consists essentially in the perception of the velocity with which the form changes.

As regards the relationship between space, *direction* and *perception*, for example, a developing strand of analysis is concerned with displacement in perceptual space (see Hubbard, this volume), and in particular with displacement perceived

in relation to the starting point (SP) and (more frequently) the vanishing point (VP) of a moving dot. Displacement occurs in various configurations, as follows:

1. The so-called *flash-lag effect* (in which flashed and moving segments, aligned in space and time, appear to be spatially displaced; Khurana, Watanabe, Nijhawan 2000; Krekelberg, Lappe 1999, which includes a review of the topic; MacKay 1958; Nijhawan 1994; Murakami 2001).
2. So-called *representational momentum* (in which the VP of a moving dot is judged to be displaced forward in the direction of motion, this being due to a memory displacement (Freyd, Finke 1984; Hubbard, Bharucha 1988).
3. A series of variations on the *Frölich effect*, which consists in a fast-moving stimulus entering a window and whose location is forward displaced (Frölich 1923: 29; Müsseler, Ascherleben 1998).

These phenomena specifically concern the fact that the influence of motion in the location of the stimulus is not restricted to the moving object alone, because even the position of stationary objects depends on mechanisms which process the movement and may even operate at a certain distance from stimuli (Whitney, Cavanagh 2000).

The fact that involved in all these phenomena is judgement of the direction and *instantaneous* position in space of a moving object (which implies being able to identify the object in a series of presentations) does not entail that these phenomena are identical. This is because they make use of different space-time markers as cues (a flash or the instantaneous appearance of an object with a frame of reference) (Whitney 2002).

This difference is not made at all clear by studies of these phenomena. More generally, experimental analysis of them is often conducted on the basis of an unwitting categorial confusion, and in particular a confusion of *spatial categorization*, which is inevitably to the detriment of the methodology used and to interpretation of the data. For example, not all experiments take account of the following factors:

1. The difference between 'actual change' and 'perceived motion' (Rensink 2002).
2. The difference between 'continuous spatio-temporal structure' and 'object', which may also be influenced by the type of display (i.e. presentation of a few static displays (Freyd, Finke 1984) versus presentation of a continuous motion (Hubbard, Bharucha 1988; Verfaillie, d'Ydewalle 1991).
3. The role of the perceptual belongingness of the field components (Benary 1924), for example how the configuration of motion in the visual field influences the location of both stationary and moving stimuli (Whitney, Cavanagh 2000).
4. The difference in magnitude among the moving elements used in experiments, which influences the velocity perceived (Runeson 1974; Bozzi, Bressan 1987).

5. The presence of colour, which influences the perception of motion (Nijhawan 1997).
6. The differing speeds of the elements (Müsseler, Ascherleben 1998).
7. The presence or otherwise of reference grids as in the original *Frölich effect*, where it alters the data through the occlusion effect that it produces (Kirschfeld, Kammer 1999).
8. The structural difference among phenomena like the *flash-lag effect* (displacement between spatio-temporally aligned flashing and moving stimuli), or the Frölich effect (a fast-moving line entering a window).
9. The differing cognitive processes involved (early mechanism vs. memory or higher-order processes).
10. The different role and effect of attention (attention shift).
11. The connection among the physical motion of an object, the perceived motion, the motion of other objects in the field, and its perceived spatial location (Whitney 2002).
12. The specific spatio-temporal structure involved in these types of events, which displays aspects of anticipation (for instance in the direction of the perceived motion).
13. The nature of the space and time involved at the phenomenal level (Albertazzi 2002a, b) both because of the influence that the *temporal coding* of a moving object has on the perceived position of an object, and which gives rise to a time delay in a spatial visualization (the explanation usually given for the flash-lag effect), and because of the importance that may be assumed by simply *where* the object is perceived, independently of the temporal mechanism (Whitney 2002) explanation for the flag-lag effect).

It is instead necessary to distinguish the various factors at work because, for example, the presence or otherwise of a grid or a window in the Frölich effect is a stable frame of reference which helps memorization of the perceived position; and this suggests a minor role of memory in the genesis of effects.

Finally, and even more importantly, it should be borne in mind that some interpretations of the phenomenon centre on psychophysical aspects (stimuli, fixation point, etc.), some on neural aspects (neural delay), and others on strictly phenomenal aspects (perceived displacements). These last may depend on various interconnected field factors, for instance the number and type (stationary/moving) of objects, or objects of differing brightness (Hess 1904).

Failure to take account of all these factors involved in the *organization of phenomenal space* gives rise to:

1. Over-emphasis of the importance of some factors with respect to others: for instance, attention (Baldo, Klein 1995), memory (Hubbard, Bharucha

- 1988), the latencies due to temporal integration (De Valois, De Valois 1991; Krekelberg, Lappe 1999; Eagleman, Sejnowski 2000), etc.
2. The equating of structurally different phenomena (*flash-lag effect*, which consists of two different types of ‘object’, one a *stationary* and *flashed* stimulus, the other a *moving stimulus*) and the *Frölich effect* (which again consists of two different types of ‘object’, one a *moving stimulus*, the other *stationary but not flashed* (the window) and which *occludes* the former) (see also Whitney, Cavanagh 2000).
 3. The use of radically different epistemological explanations (internalized mental mechanics (Hubbard, Bharucha 1988), based velocity extrapolation of a stimulus in direction of the motion (Najhawan 1994; Kurana et al. 2000), early visual mechanism (Actis Grosso, Stucchi 2003), attention shifts (Baldo, Klein 1995), integration/interpolation of the moving object’s trajectory (Eagleman, Sejnowski 2000), anticipatory retinal responses (Berry, Brivanlou, Jordan, Meister 1999), re-entrant mechanism (Whitney, Cavanagh 2000), etc.
 4. Confusion among different kinds of spatial displacement, such as the flash-lag effect, which concerns two ‘objects’, and the kind which instead concerns the boundaries of ‘one single event’, i.e. SP and VP of a movement (Actis Grosso, Stucchi 2003).
 5. The incongruence of the experimental results (for example, the forward or backward displacement of the SP and VP of the moving element) (Frölich 1923; Hubbard, Motes 2002).

Defining the conceptual categories which concern spatial perception is therefore important in regard to:

1. The research methodology to adopt.
2. The interpretation of the experimental data.

The Gestalt legacy

Some of these categories were developed by *Gestaltpsychologie*, which emphasised that perceptual processing comprises *physical*, *neural* and *phenomenal* aspects. In other words, different categorial apparatuses regulate these various aspects, with the consequence that the instruments used for their analysis affect the correctness of the inquiry. This raises the obvious problem of determining the laws of dependence among the principles regulating the various aspects: for example, the nature of the psychophysical and neural correlates of the phenomenal aspects.

In particular, the Gestalt legacy consists of:

1. Specification of the principles which regulate the phenomenal (qualitative) level of perception.
2. Demonstration that analysis should begin at that level.

Analysis conducted on these principles has shown that perceptual space has a nature, primitives and rules of organization which differ from those of physical space.

Some of the categories concerning perceived movement at the phenomenal level are the following:

1. Perceptual belongingness

The principle of perceptual belongingness is evidenced in particular by assimilation phenomena (see chromatic assimilation) (Benary 1924; Fuchs 1923; Musatti 1953; Kanizsa 1980). Analysis has revealed the presence of global processes of unification (belongingness) as opposed to segregation due to relations which depend on local factors of summative type. The principle of perceptual belongingness is part of a general tendency towards *maximum homogeneity* (Musatti 1953; Albertazzi 2004), so that perceived as units are those elements which varyingly obey the laws of field organization (proximity, similarity, common fate, etc.).

However, it has been found that phenomena of pre-attentive grouping (Treisman 1996) do not suffice to explain the perception of visual objects. Some features of objects do not occur naturally but are due to the (binding) action of the attention. Still to be analysed is the relationship among the *levels of organization* that generate vision of objects and their neurophysiological counterparts (Agostini, Galmonte 2002).

2. Change

In vision, the perception of *change* does not coincide with the perception of physical motion. Change is in fact a particular aspect of the perception of movement. Since the classic Gestalt analyses (Wertheimer, Ternus, Duncker, Lindemann, Michotte), the difference between physical motion and change has been recently repressed by studies on attentive processes (Rensink 2002; Mack, Rock 1998; O'Reagan, Noë 2001). The results demonstrate that:

1. A single perceived entity has a spatio-temporal continuity of structure.
2. In order to see change a focused attention is necessary (Rensink, O'Reagan, Clark 1997).

These studies have shown in particular that perceiving change involves various features, viz.:

1. Realization that something exists.
2. Identification of what it is that changes.

3. Perception of a changing spatio-structural structure.
4. Location of where this is happening.
5. Distinction between the process of apprehension and the content apprehended.

The change over time of an enduring structure (or of a spatio-temporal continuity) concerns both particles and complex objects. Rensink has analysed the continuity of these structures (due to contact together with the simultaneity of the parts) and their change of position over time relatively to location with respect to their point of departure and arrival. Rensink has also shown that the change of an entity is identified by two different types of description which respectively concern:

1. Description of variations with respect to *place* (i.e. motion), and
2. Description of variations with respect to a *structure* (i.e. change).

Although the two procedures are obviously correlated, they can be analysed separately.

Put in these terms, however, understanding the phenomenon of perceived movement encounters two main problems.

The first is:

1. The dependence/independence relation between the two descriptions (of motion and change); that is, the ontological dependence/independence relation between '*motion of an object*' and '*change in a continuous spatio-temporal structure*'.
2. Distinction among the various cognitive levels involved in the identification process.

The second problem, which Rensink himself has stressed, specifically concerns *change*, which may be of two types:

1. Dynamic, present, ongoing.
2. Completed, past, ex post.

Dynamic change is the ongoing change expressed linguistically by the present participle. It involves:

1. Perception of the change in itself.
2. The external spatio-temporal continuity relative to the spatio-temporal continuity of the inner representation, also in the presence of brief hiatuses.

From this point of view, *dynamic change is the visual event that deploys in the time of presentness* (the base format of presentation: Albertazzi 1999).

Completed change is instead expressed by the past participle. It concerns change:

1. Which occurs at a particular moment (for example during a momentary occlusion).
2. Which does not produce any phenomenal perception of change.
3. Which is perceived as having occurred in the past and is therefore inferred.

From this point of view, *completed change is a property of the representation in the memory and therefore concerns a higher cognitive level* (the format of representation: Albertazzi 2001a).

A distinction useful for defining the perception of movement is therefore that between:

1. Perception of change as such, i.e. perception of *temporal change in an ongoing structure* (which concerns a dynamic type of temporality), and
2. Perception of difference, i.e. noticing that lack of similarity in the properties of two already-completed structures (which are therefore considered atemporally).

The *perception of change* refers to the *same structure at different moments in time*; that is, it refers to:

1. Dynamic change and spatio-temporal continuity of both
 - 1.1 the external entity and
 - 1.2 the inner representation
2. The simultaneous presence of successive phases of the ongoing change (Benussi 1925; Albertazzi 1999).

The perception of difference instead involves comparison of two structures for which:

1. Their similarity is inferred by atemporal comparison between structures which may or may not exist simultaneously.
2. Their *relevant properties* are extracted at a certain point and *compared*.

In regard to the concept of *difference* used in psychophysics, Meinong noted (1889) that two different concepts are sometimes used univocally, viz.:

1. The concept of *Unterschied*
This is the metric difference between homogeneous divisible magnitudes (the result is the same if, for example, a segment is divided into several parts)
2. The concept of *Verschiedenheit*
This is the qualitative difference between magnitudes that cannot be divided

into parts homogeneous with the whole (for which there is real and proper unit of measurement), as in the case of phenomenal colour.

This further distinction shows that diverse cognitive processes are involved in the perception of movement, i.e. in the perception of:

1. Ongoing change, or dynamic change.
2. Completed change.
3. Quantitative and qualitative difference.

All these processes have to be taken into account in the analysis of the characteristics of visual space.

Corporeal schema

At the level of spatial perception and in particular relatively to peripersonal space, distal events are incorporated into what can be called the *corporeal schema*. The boundaries of this space are not rigidly Euclidean; rather they are characterized by a plasticity and interconnection among the various factors at work, so that as far as the specific case of distance is concerned, this appears to be one of the many non-detachable 'indexes' of the whole field.

The already mentioned examples concerning distance, position and direction of objects in space perception already show the complexity and the general anisotropy of visual space. If we then analyse the nature of the individual factors involved, we see that these are:

1. Non-homogeneous factors.
2. Factors which obey different laws of organization.
3. Factors with different weights.
4. Factors which coexist and are interdependent.

In particular:

1. None of these factors gives metrical information in a strict sense;
2. Diverse factors may refer to different 'spaces'.

Prima facie, some factors in vision appear to be external conditionings on some space: eye movements for example. Others appear to be internally constitutive of it, as in the case of distance, which acts as a component of figural objects or of perceived movement, and therefore as a non-separable parameter of the entire percept. Yet, however intuitive 'externalness' and 'internalness' seem to be, they are properties that require explanation if they are to be used properly, given the plurality of the 'spaces' involved. From the Gestalt point of view, the distinction

between inner and outer is a distinction ‘internal’ to the field itself. For example, the unilateral function of contour shows a segregation towards the inner, and in ambiguous figures the same part can play the role of external or internal, however never simultaneously and so on. At first sight it seems possible to speak of *different gradients* of externality and internality in the space of visual objects.

The intrinsic complexity of perceptual spaces therefore raises a series of questions, for example:

1. What are the *specific properties* of perceptual spaces, their *differences* and their relations?
2. What is the *role of the perceiver* in these same spaces?

In fact, the perceiver clearly plays a role as the origin of the system of spatial coordinates and thus determines the fundamental directions of space itself (so-called projection space). In this sense, Koffka described the Ego as a region segregated and functionally different from others, or as a spatio-temporal structure embedded in a field.

Pictorial space adds further complexity to analysis of visual space and exhibits a further commitment by the perceiver to the construction of the objects of vision. Pictorial space, in fact, has properties similar to those of optical space: boundaries, depth, colour, brightness, distance, and so on, which are only actualized or imagined (“looked in”) by the perceiver (Koenderink, van Doorn 2003), in that it has no physical support except the medium of the pictorial representation. It was for this reason that Arnheim called a picture an *analogue of perception* (image) *in a medium*, given that a work of art necessarily requires a material bearer, a support which enables it to take concrete form (canvas, marble, copper, film, manuscript, book, journal, CD-Rom, file, videotape, etc.) (Arnheim 1954). Of course, the choice of medium, whether a picture or a statue, entails specificities in the relative pictorial space: a statue has an *actuality* from the objectual point of view which is very different from that of a painted portrait or a photograph (Hildebrand 1945). The activation of pictorial space by the perceiver in different cases therefore follows at least partially different rules of construction.

Also the analysis of pictorial space reveals the presence of different ‘spaces’ embedded in each other and active, often simultaneously, at different representational levels. Consequently, even the concept of ‘optical information’ as a basic element of visual/spatial information is enormously more difficult to define in the case of the *visibility* and *corporeality* of pictorial objects. As regards pictorial space in particular, the question arises as to whether pictorial *objects* are types of perceptive objects different from natural ones: that is, whether they are to be considered simply as a variety of *seen* objects or whether they are already *thought* objects.

All in all, these considerations would suggest, as Th. Lipps maintained, that aesthetics is an essentially psychological discipline or a branch of applied psychology (Lipps 1879; Arnheim 1954, Chapter 9).

Theories on these various matters have analysed aspects concerning the nature of stimuli, the workings of the neurons, sensory response, the eminently qualitative nature of phenomenal vision, or aesthetic yield. But establishing correlations among events at these various levels sometimes obscures the fact that these levels are constantly in dependence relations and that, in epistemology, they are sometimes made to collapse into each other, with a tendency towards reductionism to the physical/neuronal level (see below, §2). Also testifying to this situation is a widespread terminological and conceptual synonymy in reference to spaces; a synonymy which also concerns such key concepts as information, representation, picture, image, depth, position, place, form, brightness, outline, and visual quality.

Köhler wrote in this regard:

If someone states that things seen must first be experienced as if they were in the brain, he has not realized that the first part of his statement refers to the visual field as a fact of experience, whilst in the second part, where he uses the expression ‘the brain’, he is speaking of a physical object in physical space. This means that he expects to see parts of visual space localized in relation to parts of physical space, and this notion is entirely impossible. (Köhler 1947: 213. Emphasis mine)

In other words, much of the structure present in the brain does not become conscious information: a mind is necessary.

It seems that, then as now, the general discussion suffers from some structural shortcomings. It lacks, for example:

1. A categorical classification of the differences between the physical, the neuronal and the properly psychic (mental) marking the onset of perceptions.
2. A distinction between the objectual quality of perceived objects and their conditions of perceptibility (for example, the assimilative conditions in stereokinesis, the role of subjective integrations, etc.) (Benussi 1922/23; Musatti 1964).
3. A univocal terminology (representation, perception, mind, image, picture, visual space, information, shape, etc.).
4. Feedback from established computational models on the theory of vision.

My thesis is that:

Only the framework provided by a properly developed theory of levels of reality can handle the complexity of perception space.

A second step is the complementary adoption of:

A dynamic and processualistic theory of reality which explains continuity within the levels of reality.

Both theses are based on a preliminary analysis of the concept of ‘representation’.

Representation

The concept of ‘representation’ is one of the most ambiguous in the cognitive sciences, and as a consequence it is ambiguous in the theory of vision as well. Its ambiguity is due to the fact that on the one hand the term refers to the taken-for-granted *external independently existing object* to which our *acts* are directed, and on the other to the more or less accurate *image* of the object which exists *internally* in our mind (and/or brain, according to the disciplinary background of the researcher). Apart from the exponents of an extreme constructivism (Glaserfeld 1984; Watzlawicz 1984. See also Maturana, Varela 1987), no one seems to question that the ontological weight of existence resides entirely in three-dimensional objects existing in the so-called ‘external world’, or the position-space of physics and that they are endowed with the properties that the latter ascribes to them (Marr 1982). Also positions close to a phenomenological psychophysics, for example, share the idea that perception consists in detecting *objective properties* of the world like the illumination, brightness, distance, direction and so on of physical objects (Gibson 1979; Shepard 1994; Palmer 1999; Kubovy et al. 2002. For a critique of the various positions taken up by theorists of perception see Vishwanath 2005, and this volume. See also Hatfield 2003). Generic ecological optics, for example, coincide with a subset of physics, assuming the point of view of a ‘standard’ observer (see Koenderink, van Doorn 2003:243).

From this point of view, the perceptual system informs us about elementary physical quantities of whatever type, intended as energy of sounds, intensity and wavelength of light. In short, these theories take identification of the *phenomenal object* with the *physical object* for granted and do not consider it at all problematic (see for criticism Mausfeld 2002, 2003).

This position has been taken up and strengthened by a functionalist and computational theory which views perception as the *recovery of physical world structure* from sensory structure by means of input-based computational processes, so that, for example, in inverse optics we go from image to the true 3D layout of physical objects in a scene (Poggio 1990). According to this theory, information is already stored in the external world, and the task of representation is simply to extract its invariants.

In general, the predominant idea is that representation, or the onset of the psychological level of consciousness within the primary cognitive processes, is *internal*. Only recently has the proposal been made of an ‘extended mind’, although this has considerable categorial weaknesses, principally the fact that it entirely lacks a dynamic and stratified view of reality. From this perspective, the objects of the mind – be they tables, marbles, hourglasses, Knights of the Round Table, pens, nations, wars, quanta, neutrinos, Shakespearean Juliets, abstract figures and personal memories – are all embedded in an ‘extended space’ or wide ware dependent

on various cognitive technologies (Chalmers 1996; Clark 1991). This conception defends an active externalism, based on the role of the environment in driving cognitive processes (for a criticism, see Marconi 2005). However, to extend the mind to the world, assuming a sort of good store situated outside of the brain, does not solve the problem of how mind arises. From this point of view, moreover, in this framework it will never be possible to analyse the complexity of visual space.

The best-known current views in the field of representation theory can be grouped into two general positions: one relative to neuronal research, and the other relative to the symbolic approach developed in AI.

The former position consists in the belief that representation originates in *cortical maps*. It divides into various proposals: the role of coherent oscillations (40–70 Hz) in connecting together percepts pertaining to a single perceptive moment (Singer, Gray 1995); the view that cognitive processes originate as a quantum process in the neuronal microtubes (Penrose 1994); reentrant signalling across cortical maps (Edelman 1989), or the neuronal correlates of consciousness (Chalmers 1996). Despite recent and indubitably significant advances in the neurosciences, this position has been criticised on the grounds first put forward by Köhler that even if we assume that every perceptive event has a neuronal correlate, it does not follow that the *entire dynamic* of neuronal processes is represented by corresponding phenomenal facts.

As regards the second position, which now seems to be losing authoritative-ness, the early versions of Fodor's 'language of thought' and Pylyshyn's 'computational approach to human information processing' can be read as examples of linguistic-oriented representation theory. In some ways, this conception is an heritage of the 20th Century philosophy of mind, bearing traces of physicalism and behaviourism. Analytic philosophy, the main philosophical trend during the last century is, in fact, but a form of behaviourism.

The fact that the concept of representation is posited in such disparate terms raises several sub-issues, for example:

1. Are there *primitive aspects* of representing connected to a type of *concrete* presentation which differentiate its format with respect to *abstract*, mnemonic or symbolic representations? And precisely what should 'concrete' mean in this respect?
2. Is it possible to draw a demarcation line, at least *descriptively*, between the different aspects of the representation, i.e., those relative to the psychophysics of stimuli, those relative to neuronal activity, and those relative to phenomenal appearances in the various perceptive fields?
3. What difference is there between the *object* and the *content* of a representation?

As regards the third point, in fact, 'having something *as* an object', is not the same thing as 'having something *for* an object,' as the original theory of intentional-

ity stressed (Brentano 1995). In other words, considering representation from the point of view of the ‘internal correlate *of* the act’ (content) is not the same thing as considering it from the point of view of ‘the object *presented by* the act’ (object) – whatever is meant by the term ‘object’. The distinction is of great importance in analysis of pictorial space, for example (see below, §3).

Within a more general theory of representation, therefore, as a first step I propose to distinguish a *base format* in ongoing perception. This I shall call *presentation*, whether objects or contents are concerned. Characteristics of presentation are:

1. A presentation denotes the *inner* and *actual* aspects of the general theory of cognitive processes.
2. A presentation is *not* understood as a *representation of the external world*: that is, the connection between the presentation and its triggers (the stimuli) is not primarily taken into account by this particular type of structure.
3. A presentation as such *presents, exhibits the structure of what appears in the now* (the time of presentation) bracketing off the physical and the neuronal levels of the objects of vision.³

From this point of view, information is not deposited once and for all in the external world. Rather, it is the presentation as a specific spatio-temporal structure that confers meaning on the scene towards which it is directed. This, in short, is also the sense of *intentional act* and *directedness towards an object* in Brentano’s theory (Brentano 1995). From this also derives the intrinsic *ambiguity* or plurivocity of perceptive, or more generally cognitive, situations, which are intrinsically tied to the point of view, the mode of observation, and the laws of perceptive organization of which that structure is a constitutive component. In other words the bearer of the presentation is a non-independent part of the whole.

A presentation, moreover, is the *point of departure* for analysis of experience. This entails abandonment of two generally held assumptions in the contemporary theory of perception, i.e.:

1. There exist unquestionable and well-defined 3D objects in the external physical world (for example, tables, chairs, glasses) which emit stimuli that we *represent* by means of neuronal processes. These processes, however, bear no similarity to either the stimuli or our representations (images).
2. The direction of representation is always and unequivocally from *outside to inside*, so that the organization of perceptive experience is an almost entirely parasitic on stimuli.

Köhler was the first to raise serious doubts concerning this conception of classical psychophysics, and the model was subsequently challenged by Gestalt psychology,

with its view of a *functional* correspondence between the *spatio-temporal structure* of perceptual events and their *modes of behaviour* in the field.

The phenomenological approach to the theory of perception, in fact, starts from an assumption profoundly different from that of classical psychophysics. Aside from individual positions, which may differ markedly (consider the differences within Gestalt theory between the theories of Ehrenfels, Benussi and Koffka, or of Köhler or Brunswick), a phenomenological approach to the theory of perception highlights the existence of a *subjective aspect* which is constitutive of the construction of empirical reality.

As Michotte writes, it is through *events* – which must be understood in a very broad sense as sensory, intellectual, affective, etc. – not through stimuli, that we come to know not only objects, processes or group of objects or processes, but also complex spatial-temporal situations. These events are the basis of *all* the knowledge that we acquire about the physical world (Michotte 1950a). In other words, in this framework the directional ‘arrow’ of inquiry goes in the opposite direction, from the perceptual events (of qualitative nature) to stimuli (of quantitative/metric nature), and consequently perception is not viewed as the *mere transformation* (and *re-presentation*) of a stimulus into something else.

Of course, distinguishing a presentative level of our experience as original (and coinciding with phenomenal events), as I do, *is not at all to question the existence of underlying levels or of dependence and interconnection relations among those levels.*

Levels

The theories of the levels of reality available in contemporary ontology are unequal to the task because of their conceptual ingenuousness. According to a naïve theory of levels, in fact:

1. A level is a large collection of units.
2. The level structure is serial (one level after another).
3. The level structure is pyramidal (lower-levels are wider) (On the theory of levels see Poli 2001).

There are various problems with this approach. It assumes a *linear* order (for example, material and psychological beings, or social and psychological beings, simply depend on one another). It is difficult to determine dependency connections. And the internal segmentations of the individual levels assume different patterns. This approach is consequently intrinsically reductionist, because:

1. Classical physics is considered to be the basic level, so that all the others are non-basic.

2. Each level is composed of objects pertaining to the level below it.
3. For every causal statement true at some non-basic level, there is a corresponding causal statement that is true at the basic level (Oppenheim, Putnam 1958: 3–36).

However, the main point is that, even in cases where reference is made to a ‘theory of levels’, most contemporary approaches treat questions of emergent evolution, supervenience, complexity and non-linearity in terms of a *theory of descriptions* (Quine’s heritage), according to which reality can be segmented in any way that is conventionally and pragmatically useful. This is an epistemological distinction which may be useful in many cases but should not be generalized. Some differences, in fact, like that among the levels of reality, are instead *ontological* in nature, and they are characterized by different types of causality or causal dependence. Causality in physics, for example, is very different from the figural relation between the perceiver and certain structures in his or her environment: Köhler called this ‘meaningful connection’ (*verständlicher Zusammenhang*) and emphasised its functional nature, and Michotte has analysed its rules of organization.⁴ Whatever the case may be, it should be borne in mind that there exist *lines of tendency*, ‘natural nervatures’ of being as Plato called them, which ontologically segment reality and which cannot be eliminated or changed at will.⁵

Here, in order to specify the difference with a general theory of levels, I shall use *level* as a general term, *stratum* to refer to the realms of being characterized by categorically diverse groups (for example, the physical or the psychological stratum), and *layer* to refer to the segmentation internal to each stratum.⁶

The first point to be considered is that relations among levels and sub-levels may be of different nature. A good distinction in this respect is that between *building-above* (*Überbauung*) and *overforming* (*Überformung*) relations.⁷

Strata or levels of reality are connected by building-above relations, i.e. by the birth of a *new categorical series*. For example, the group of categories that are needed to analyse the phenomena of the psychological stratum – universal categories apart – is different from the group of categories needed to analyse the social or the material stratum of reality.

Overforming relations are weaker kind of relations, and they are generally used to analyse the internal organization of strata.

From a categorical viewpoint we have *universal* categories (for example, space, time, part, whole) regarding all the levels, and *level* categories, pertaining to the different strata, which are new categorical series.

Of the three fundamental strata of reality – which, following Poli’s distinction are physical, psychological and social⁸ – the material stratum is the one best known and most widely investigated by science. Its physical, chemical and biological components are usually distinguished, and they can be broken down further.

For example, biology can be distinguished into genetics, cytology, physiology, ethology, ecology, and so on (Poli 1998, 2001:268–69).

What distinguishes the psychological stratum from the material stratum is an interruption *in the material categorical series and the onset of new ones*. In fact, in the psychological, and in the social stratum as well, we find more complex types of *over-forming* relations instantiated. The categories of the psychological stratum therefore manifest *categorical independence* (newness) and *existential dependence*, because in order to exist requires a physical bearer, the perceiver. This is exactly the building-above relation between strata.

Time and space (or the part/whole relation of objects) *differ* in the various ontological strata. For example, it is obvious that psychological time differs from physical time, but there is no theory that adequately explains the difference; rather, metaphors are the rule.⁹ The same applies to the space of the objects of the phenomenal visual field, which are often analysed only in terms of their psychophysical or neuronal dependence, owing to a *horror psychologiae* of Enlightenment origins.

More specifically, embedded in the psychological stratum – made up of specific layers such as the primary and secondary cognitive processes – are diverse space-times, or diverse families of space-time. The mental spaces and times of actual perception, of memory, of deductive inferences, or of artistic creativity, for example, manifest granularity and complexity and have different components which are not reducible one to another.

The questions to be addressed, therefore, are whether the stratum of perception (and *visual* perception in particular) has:

1. Building-above relations with respect to the physical/neuronal stratum.
2. Different ‘spaces’ embedded in the various layers.
3. Building-above phenomena (not only *over-forming* ones) also within its layers.

As regards the third point, the question is whether relations of categorical ‘novelty’ concern not only the psychological stratum but also its various layers, and whether as a consequence the objects of these layers are located in different ‘psychic/mental spaces’. For example, if natural perceptual objects and pictorial objects not only relate to different spaces, *but in what do these spaces consist*; and if pictorial objects are built on a sort of continuity with natural perceptual ones. The discovery of forms of continuity between natural and pictorial objects, moreover, would lead to integration of the *descriptive* content of the theory of levels – which essentially envisages discontinuities – with a theory that takes account of the dynamics internal to the levels and of the phenomena involved (see below, §3).

Identifying relations of categorical novelty is essential for analysis of the structural diversity of the perceptual visual field as identified by Kanizsa in his classic studies (Kanizsa 1991).

Kanizsa distinguished between sensory, amodal and mental modalities in perception, and considered the logic of perception to be different from the logic of thought, so that 'seeing' and 'thinking' have different types of representation and different laws of organization.

Experimental analysis conducted according to this heuristic, which separates seeing and thinking, shows that different – though often co-present – forms of completion operate in the two different types of representation, but it also shows that the *boundaries* between seeing and thinking are very difficult to establish. Still unresolved, therefore, is the question of whether there is a separation, with a possible categorial building-above differentiation, between the layers of seeing (primary psychic process) and thinking (secondary psychic process) of the psychological stratum, or whether, as Arnheim has argued, they are the two poles of the same continuity (Arnheim 1969).

In other words, the question is whether there is continuity between the different types of subjective integration (bottom-up and top-down) or whether they are two separate classes of completion related to separate classes of objects and *over-forming* categories.

From this point of view, too, pictorial objects constitute a very interesting field of study. The continuity or discontinuity between a 'perceptive presence' and a 'mental presence' (for the distinction see Benussi 1925; Kanizsa 1991), and between these and memory, in fact, concern both the *seeing of natural objects* and the *seeing of pictorial objects*.

Objects

The analysis of the visual objects bring us to the second methodological and ontological option mentioned at the outset and which, after adoption of a descriptive theory based on the levels of reality, gives rise to a dynamic and processualistic account of the reality observed. The difference is the following.

The ontological theory of categories *distinguishes* and *describes* types of entities (objects). The assumption, however, is that these entities are not totally independent from one another, but that these entities (types of both substances and accidents) are connected by a network of dependencies in the different levels of reality.

The theory of processes, instead, addresses the problem of reality from the point of view of *continuity*, not of discreteness. Specifically, it concerns:

1. The processes of transformation occurring within layers of reality.
2. The transformation beginning in a layer and ending in another, passing through intermediate ones.

At the moment the question whether continuity occurs not only within layers but also among strata is not addressable, due to the lack of both data and a theoretical framework. From an ontological point of view, however, it seems more plausible to maintain a *categorical novelty among strata* in order to justify the presence of new emergent categories.

The question of *continuity within layers* of a stratum is particularly relevant in the case of vision and the *objects* of visual field. An important concern of the analysis of visual space is the structure of ongoing perception (corresponding to my concept of ‘presentation’), which configures itself as a sort of ‘actual performance’. *Seeing* an object is to see it *in its unfolding*. It is not simply to have an image ‘represented’ or recorded in the brain of something already constituted with all its properties in the physical world. It is instead to participate in its processualistic dynamic, to contribute at least in part to its construction. The spatio-temporal subjective structure of the actual presentation – whatever is meant by space-time at this level¹⁰ – is a constitutive *part* of the object itself. This is at bottom the idea of ‘intentional object’: an intrinsically subjective object (at the primary and/or secondary level) emerging from the structures of acts of presentation like seeing, hearing, feeling, thinking, remembering, imagining, and so on (Brentano 1995; Albertazzi 2005, Chapters 3, 4).

Visual objects have a complex structure: for example, there is almost always a ‘double representation’ in the phenomenal field produced by the figure/ground articulation. More generally, phenomenal objects are characterized by multistability, occlusion, anomalous contour, stereokinesis, masking, and transparency. Consequently, more than objects identifiable according to well-defined classes of properties, there are appearances in the perceptual field with strong dynamic and contextual connotations.

Moreover, the ‘stuff’ of which these forms are made up is not reducible to purely physical dimensions. Consider the brightness and the whiteness of Kanizsa’s amodal triangle, which is entirely non-existent from the point of view of stimuli, the ‘incorporeal’ appearance of shadows (Metzger), or conversely the corporeality of stereokinetic objects (Musatti), the transparency of certain surfaces, the continuous transformation of colour on a surface, and colour illusions (Da Pos 1997).

Among the ‘intentional’ characteristics of the objects of a visual perception are the following:

1. Specific characteristics of their visibility, during the act of seeing.
2. A complex construction of their phenomenal appearance.
3. An intrinsic process-structure.
4. A dependence on the act of seeing.
5. A continuity/discontinuity in the visual space.

6. A specific kind of identity, which is not of logical type (Michotte et al. 1962).
7. An expressive value.

Assuming this point of view means to have a very different conception of 'object'. The objects of the phenomenal visual space, in fact, configure themselves as *forms*, and as *relational structures*.

According to the Gestalt laws:

1. Forms are *cohesive* wholes (i.e. not mere aggregates) where the *position* of the parts plays an eminent role.
2. The whole is *more* than the sum of its parts.
3. The whole *emerges* in a specific relation with its parts.
4. The whole has a *boundary* which separates it from the ground, and may have internal boundaries of different type relative to its components.

Forms are characterized by a *part/whole relation* in which of prime importance are the position, salience, orientation, ambiguity, similarity and connectedness of the parts.

This basic relation, however, is problematic because it is intrinsically dynamic and occurring in multiple and multifarious ways. Moreover, not only do there exist different types of forms but also different types of parts.¹¹ In this respect one can first distinguish between (Henry 1991; Poli 2001):

1. x-parts = simply parts, or parts which may become such (i.e. they are not actually); these are *antecedent* parts necessary for a whole to exist (for example, the design of an artefact or its material).
2. Parts-of = x-parts when they are incorporated in other objects; these are *contemporaneous* parts in reciprocal determination of the whole (hand in a body, legs in a chair, parts of a rectangle).
3. *Subsequent* parts = those which *arise on completion of the whole*, or which require a preceding *phase* of existence of the whole (as the hypothesis of proto-objects in the study of attention would support) (Rensink 2002).

Secondly, parts can be distinguished in *independent* parts (pieces) and *non-independent* parts (moments) (Husserl 1891: Ch. 11, §6ff. and 1970, Third Logical Investigation).

1. Independent parts are parts which are removable from the whole (in inanimate objects, for example, a leg of a chair, in physical processes what can be placed in another temporal moment).
2. Non-independent parts are parts which can be conceptually or phenomenally distinguished from the whole but cannot be phenomenally separated (for example, colour and extension).

Also in the case of forms which appear in the visual field it is possible to distinguish their parts as independent parts from non-independent parts of the whole.

Examples of independent parts are the two parts of a rectangle divided by a line, where the same elements may exist and are experienced in two different modes, i.e. in grouping and in isolation.

Typical examples of non-independent parts are:

1. Colour and surface (parts which are not separable, named 'essential').
2. The partial intuitions of the whole (for example, the configuration of the black squares of a chessboard vs. the white ones; or the two parts of a rectangle divided by an *oblique* line: in this case the difference is given by the fact that the superior and inferior sides of the rectangle are continuous straight lines, while the sides of the two parts pertain to irregular figures).
3. The inner relations connecting the various parts of the whole (Gestalt qualities proper).

The situation is further complicated by the fact that forms manifest *forces* which organize themselves around centres in accordance with specific laws: for example, organizing themselves around the physical centre of gravity are other perceptual centres constituted by often 'invisible' (though perceptively active) force lines which act in the configuration and apprehension of the visible whole (Arnheim 1954:31–32). The velocity, direction, boundary and weight of phenomenal visual objects follow the dynamic which arises among verticality, horizontality, colour, brightness and other components of the field – as Arnheim's analyses have well shown both for natural and pictorial perception (Arnheim 1988, Chapter 1; Leyton 1992; Albertazzi 2002b). The part/whole relation can therefore be considered from this further point of view as well.

Contents

Another distinction that an analysis of visual objects must draw is that between *objects* and *contents*. The distinction can be understood by considering the fact that a general theory of objects (ontology) also deals with immanent, transcendent, negative, impossible, incomplete, non-existent, past or future objects, which however we are able to represent to ourselves and to address according to their *content*. We represent to ourselves, in fact, the winged horse, the Punic Wars, God, the concept of impossibility, the square circle, and so on.¹² In other words, *contents are specific types of objects*, and in perception (natural and pictorial as well) they play a relevant role.

As regards visual phenomenal objects in natural perception, for example, *objects and contents coincide*, as do the times of their ongoing presentation (Meinong

1899). In pictorial space, by contrast, *objects are presented according to their content depicted*: indeed, pictorial space can be defined as ‘the space in which the pictorial content unfolds’ (the expression is from Koenderink, van Doorn 2003: 248).

Considered from this point of view, an object of pictorial representation manifests diverse ontological levels, as the analyses by Hartmann and Ingarden have shown (Hartmann 1950, 1952; Ingarden 1962, 1968). Hartmann, for example, distinguishes two layers in the aesthetic object: the *foreground* layer, which comprises the physical dimensions of the object, and the *background* layer, which is the layer of the *content* embedded in the foreground layer. This background layer, however, exists only in relation to the observer who grasps the content. Ontologically speaking, therefore, the aesthetic object has two layers, and the foreground layer imposes constraints on the background layer. The background layer, however, according to the overall structure of its content, has many layers within it, depending on the type of object, for example whether this is a literary work or a painting (Hartmann 1950).

If we consider a pictorial object like a portrait, for example, its background layer has several layers, i.e.:

1. The 3D space in which the subject of the portrait and some elements of the setting appear.
2. The movement of the subject’s apparent corporeality.
3. The subject’s character.
4. His/her individual idea, or the idea that the person portrayed has of him/herself.
5. The symbolic (universal) content manifested by the portrait.

Ingarden concerned himself in particular with aesthetic objects of literary type. But there is a component of his analysis which can be generalized to all kinds of aesthetic objects (and, I believe, to all types of presented objects): the concept of *points of indeterminacy* (*Unbestimmtheitsstellen*). Points of indeterminacy are points where the aesthetic object, which in principle can be described in terms of a number of finite properties, does not furnish details which must be filled in – to a certain extent – by the perceiver (Ingarden 1937, 1968). Which shows that aesthetic objects are highly ‘intentional’, that is, subjectively and cognitively supplemented.

To conclude, as regards objects of pictorial space, we may roughly distinguish among:

1. A *physical* level (seeing a canvas covered with pigments in a certain simultaneous order).
2. A *phenomenal* level (seeing a perceptive object as depicted content).
3. A *pictorial* level (seeing *in it* an object in the space of the depiction) (Koenderink, van Doorn 2003: 239).

All three levels have a feature in common – that of *seeing* – although the *mode* differs among them (on the difference among the various types of acts see Husserl 1970, First Logical Investigation).

When we see a picture, something similar happens to when we see certain so-called perceptive illusions, where a sort of split between *object* and *content(s)* occurs. In the Hering illusion, for example, the vertical lines (object) are seen as curved (content a), but with a certain effort and practice they can also be seen as parallel (content b).¹³

Analogously, the act of seeing in the perception of pictorial space gives rise to two different *states of affairs* (Benussi 1905, 1906): one which corresponds to the perceptual seeing, which has a particular object (the pictorial object), and one relative to conviction about the existential status of the object, which relates to the physical object (the pigmented canvas) and to the object thought (the depiction). In pictorial space, therefore, as in phenomenal space, several ‘spaces’ simultaneously coexist and are populated with different kinds of objects. In some cases, the pictorial object has such perceptive evidence that it may momentarily obscure the physical level of its representation.

Moreover, because the pictorial object as such is a form of reified intentional object, its analysis, by means of backwards information retrieval process, sheds further light on the nature itself of intentional objects.

In truth, analysis of pictorial objects also shows that, in general, visual objects have a different degree of *veridicality*. Consider again perceptive illusions, but more in general the plurivocity of phenomenal appearances in the various perceptive fields. The question of veridicality is meaningful if the primary level of reference is the physical one. But the majority of the objects of seeing are objects operating on the *representational structures of vision* or as field objects, i.e. they have another primary level of reference. Veridicality, for example, is almost meaningless in pictorial space, whose objects are built on second-order representational structures and are typically intentional objects (Koenderink, van Doorn 2003).

In my opinion, however, almost all appearances of phenomenal space are objects of this type, that is, not merely ‘physical’ objects. In fact, natural phenomenal objects are dynamic constructions.

Although these aspects can be mapped against the benchmark of the stimulus, they cannot be reduced to it, because they involve a *level* of subjective completion: that is, emergent components relative to the psychological level of reality (Albertazzi 2003, and this volume).

The depictive space of the mind

In many cases it has been precisely pictorial ‘models’ – with their variety both historical and stylistic – that have best displayed certain structures of phenomenal appearances, confirming the close relation between aesthetics and perception, and between the laws of perceptive organization and the laws of pictorial organization. It is as if the best artists, although they adopt different depictive styles, on ‘seeing’ visual objects – or certain of their features, sometimes more object directed, sometimes more subject oriented as in abstract art – ‘present’, and ‘achieve’ them *directly in the depiction*. Very often, therefore, the work of art has been the ‘best’ model of the complexity of perceptive organization, and of visual organization in particular, given that *it depicts aspects and/or primitives of the actual perception at different levels of subjective completion*. It is for this reason that I use the expression ‘the depictive space of the mind’ when discussing the cognitive operations involved in perception.

Consider, for example, the pictorial rendition of the difference between perceived brightness and illumination; or how the depiction conveys the perception of *surfaces* in the various pictorial styles. For example, the depiction of surfaces in Ingres’ *The Source* refers to and shapes the actual corporeal perception of a body, where the role of subjective completion is essentially modest, because in seeing the picture we get a mixed impression of volume and surface which is very similar to the natural one (Arnheim 1954, Chapter 3). Vice versa, in Boccioni’s *The Cyclist* there is a great deal of subjective completion, and the perception of the surface is rendered by simultaneous directions in the movement depicted, which refers to the genesis of the percept itself in the brief scale time of presentation (specious present). In the case of Cezanne’s *A Body*, the role of subjective completion is accentuated by means of what I have called building-above properties, emergent categorial properties. Finally, the perception of surfaces (or better ‘bodies’) in Picasso’s *A Body* is conveyed by a subjective completion based on inner spatial relations (what Arnheim calls the inner skeleton of the visual object) (Arnheim 1954, Chapter 2), and on the activation of different points of view for the different parts (Koenderink, van Doorn 2003).

At work in these various cases are mental or intentional spaces to which brain imaging seems to have little to offer because they are emerging at the pre- and representational levels of consciousness (appearances) and as such are intrinsically ‘plurivocal’, i.e. essentially cognitive (Shepard 1995; Koenderink, van Doorn 2003).

To conclude: the distinctions between presentation/representation, present/absent objects, levels, and the role played by the perceiver (as operator of the projection space) in unifying the continuity of levels, positions the discussion on the concept of visual space, and of perceptual space in general, *between ontology and psychology* (on the different types of objects see Meinong 1960. Specifically, in

present and absent objects see Benussi 1922/23:20–21, 1925, Chapter 1; Kanizsa 1980:89. On the topic see Albertazzi 1996). In particular, my thesis consists of:

1. Stressing the existence of a phenomenal level of experience characterized by categorial novelty and governed by laws of qualitative organization.
2. Considering the complexity of this level as the starting-point for scientific investigation (Michotte 1950a, b).
3. Emphasising that effort must be made to define the laws of dependence among the various levels.
4. Asserting a certain autonomy of the psychic and mental as the source of emergent structures created by perception not reducible to the material aspects of existence.
5. Realizing the laws of similarity between the mental operations acting in visual perception and the mental operations that express them artistically in a work of art.

From this point of view, Brentano might also have been right to affirm that ‘psychology, in so far as it is *descriptive*, is far more advanced than physics’. But then, a diverse conception of the science of psychology is needed, which comprises the inner complexity of its dynamics as well as its material roots: the laws of existential dependence of its categorial newness.

Notes

1. But also more recent studies, like those on ‘representational momentum’, show that perceived and remembered space undergoes considerable distortions: see Finke, Freyd, Shyi (1986); Freyd (1987); Hubbard (1999). However, these studies too lack a theory of levels of reality: see e.g. Hubbard (2004).
2. Perception of distance in pictorial space, for example, requires diverse concepts of distance applied to the different direction of the gaze. On this point see Koenderink, van Doorn (2003). On 3D space see Massironi (1998, Chapter 3).
3. The term ‘*Vorstellung*’ denotes the inner aspect of the representation related to processes of imagery in the broad sense as cognitive *capacity*. See Kant (1781 A99); Brentano (1995:97). On the topic see Albertazzi (2001a), and this volume.
4. This type of relation also concerns, for example, the aesthetic enjoyment of art: Köhler (1947:326), Michotte (1950b).
5. Plato was referring to the law of division (which is complementary to the law of unification in the dialectic) which concerns the division of the genus into its species. See *Phaedro* 266 De.
6. This distinction, which differs to some extent from Hartmann’s account, is set out in Poli (2001).
7. The distinction is made in Hartmann (1935). On Hartmann see *Axiomathes* special issue 12, 3–4, 2001, in particular Albertazzi’s contribution, which specifically refers to Kanizsa’s work.

8. Hartmann's original division was among inorganic, organic, psychic and social being. See Hartmann (1952).
9. The best analyses in this field are still those of the theoretical and applied phenomenology of the early twentieth century. See Husserl (1966); Benussi (1913); Calabresi (1930); Bonaventura (1929); Fraisse (1963). On the topic see Albertazzi (1999); Varela (1999).
10. Albertazzi (1999, 2001b). Note that by 'intentional object' is usually meant something different: the object to which the intentions (or propositional states, beliefs) are directed.
11. The best introduction to the topic of 'part' is still that provided by Aristotle in *Metaphysics* 5, 25, according to which parts can be defined according to the quantity (*poison*), according to the form (*eidōs*), and according to the whole (*holon*). These can be contemporary or antecedent to the whole.
12. This aspect has been one of the cornerstone points of Meinong's theory of objects. See Meinong (1960).
13. A very detailed analysis of the difference between object and content of presentation is set out in Twardowski (1977). See also Husserl's criticism in Husserl (1994).

References

- Actis Grosso, R., & Stucchi, N. (2003). Shifting the start: Backward mislocation of the initial position of a motion. *Journal of Experimental Psychology: Human Perception and Performance*, 29 (3), 675–691.
- Agostini, T., & Galmonte, A. (2002). A new effect of luminance gradient on achromatic simultaneous contrast. *Psychonomic Bulletin and Review*, 9 (3), 264–269.
- Albertazzi, L. (1996). A cubist state of the mind. In L. Albertazzi (Ed.), *The Philosophy of Alexius Meinong* (pp. 5–16), special issue of *Axiomathes* 7 (1–2).
- Albertazzi, L. (1999). The time of presentness. A chapter in positivistic and descriptive psychology. In S. Cattaruzza (Ed.), *Vittorio Benussi* (pp. 49–73), special issue of *Axiomathes* 10.
- Albertazzi, L. (2001a). Presentational primitives. Parts, wholes and psychophysics. In L. Albertazzi (Ed.), *The Dawn of Cognitive Science. Early European Contributors* (pp. 29–60). Dordrecht: Kluwer.
- Albertazzi, L. (2001b). The legacy of the Graz psychologists. In L. Albertazzi, D. Jacquette, & R. Poli (Eds.), *The School of Alexius Meinong* (pp. 321–345). Aldershot: Ashgate.
- Albertazzi, L. (2002a). Continua. In L. Albertazzi (Ed.), *Unfolding Perceptual Continua* (pp. 1–28). Amsterdam: Benjamins Publishing Company.
- Albertazzi, L. (2002b). Towards a neo-Aristotelian theory of continua: Elements of an empirical geometry. In L. Albertazzi (Ed.), *Unfolding Perceptual Continua* (pp. 29–79). Amsterdam: Benjamins Publishing Company.
- Albertazzi, L. (2003). From Kanizsa back to Benussi: Varieties of intentional reference. In L. Albertazzi (Ed.), *The Legacy of Kanizsa in Cognitive Science* (pp. 239–259), special issue of *Axiomathes* 13, 3–4.
- Albertazzi, L. (2004). Stereokinetic shapes and their shadows. *Perception*, 33, 1437–1452.
- Albertazzi, L. (2005). *Immanent Realism. Introduction to Brentano*. Berlin-New York: Springer.
- Arnheim, R. (1954). *Art and Visual Perception. The Psychology of the Creative Eye*. Berkeley: The Regents of the University of California.

- Arnheim, R. (1969). *Visual Thinking*. Berkeley: The Regents of the University of California.
- Arnheim, R. (1986). *New Essays on the Psychology of Art*. Berkeley: University of California Press.
- Arnheim, R. (1988). *The Power of the Center*. Berkeley: University of California Press.
- Baldo, M. V., & Klein, S. (1995). Extrapolation or attention shift? *Nature*, 378, 565–566.
- Benary, W. (1924). Beobachtungen zu einem Experiment über Helligkeitskontrast. *Psychologische Forschung*, 5, 131–142.
- Benussi, V. (1905). La natura delle cosiddette illusioni ottico-geometriche. In G. Sergi (Ed.), *Atti del V Congresso Internazionale di Psicologia* (pp. 262–267). Rome: Forzani & C.
- Benussi, V. (1906). Über Vorstellungsinadequätheit. I. Das Erfassen gestaltnmehrdeutiger Komplexe. *Zeitschrift für Psychologie*, 42, 22–25.
- Benussi, V. (1913). *Psychologie der Zeitauffassung*. Wien: Hölder.
- Benussi, V. (1922–23). *Introduzione alla psicologia sperimentale. Lezioni tenute nell'anno 1922–23*, typescript by Dr. C. Musatti, Fondo Benussi. Milan: Bicocca University.
- Benussi, V. (1925). *La suggestione e l'ipnosi come mezzi di analisi psichica reale*. Bologna: Zanichelli.
- Berry, M. J. et al. (1999). Anticipation of moving stimuli by the retina. *Nature*, 398, 334–338.
- Bonaventura, E. (1929). *Il problema psicologico del tempo*. Milan: Istituto Ed. Scientifico.
- Botvinick, M., & Cohen, I. (1998). Rubber hands 'feel' touch that the eyes. *Nature*, 391, 756.
- Bozzi, P., & Bressan, P. (1987). On the perception of launched projectiles. In *Proceedings of the Fourth International Conference on Event Perception and Action*, 49. Trieste: Dec Sistemi.
- Brentano, F. (1995). *Psychology from Empirical Standpoint*, ed. by L. McAlister. London: Routledge.
- Brunswick, E. (1956). *Perception and the Representative Design of Psychological Experiments*. Berkeley: University of California Press.
- Calabresi, R. (1930). *La determinazione del tempo psichico*. Florence: Bemporad.
- Chalmers, D. J. (1996). On the search for the correlate of consciousness. Paper presented at the Tucson II Conference: *Towards a Science of Consciousness*. Tucson, Arizona.
- Clark (1991). *Being There: Putting Brain, Body and World Together*. Cambridge, MA: MIT Press.
- Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing: The integration, relative potency, and contextual use of different information about depth. In W. Epstein, & S. Rogers (Eds.), *Handbook of Perceptual Cognition. Perception and Space Motion Vol. V*. (pp. 69–117). New York: Academic Press.
- Cytowic, R. E. (1989). *Synesthesia. A Union of the Senses*. New York: Springer.
- Da Pos, O. (1997). Colour Illusions. *AIC Color* 97 (34–41). The Color Science Association of Japan: Kyoto.
- De Valois, R. L., & De Valois, K. K. (1991). Vernier acuity with stationary moving Gabors. *Vision Research*, 31, 1619–1626.
- Duncker, K. (1929). Über induzierte Bewegung. *Psychologische Forschung*, 12, 180–259.
- Egelman, D. M., & Sejnowski, T. J. (2000). Motion integration and postdiction in visual awareness. *Science*, 287, 2036–2038.
- Edelman, G. M. (1989). *The Remembered Present*. New York: Basic Books.
- Ehrenfels, Ch. von (1890). Über Gestaltqualitäten. *Vierteljahrschrift für wissenschaftliche Philosophie*, 14, 242–292.
- Finke, R. A., Freyd, J. J., & Shyi, G. C. W. (1986). Implied velocity and acceleration induce transformations of visual memory. *Journal of Experimental Psychology: General*, 115, 175–188.
- Fraisse, P. (1963). *The Logic of Time*. London: Methuen.
- Freyd, J. J. (1987). Dynamic mental representations. *Psychological Review*, 94, 427–438.

- Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10, 126–132.
- Frölich, F. W. (1923). Über die Messung der Empfindungszeit. *Zeitschrift für Sinnespsychologie*, 54, 58–78.
- Fuchs, W. (1923). Experimentelle Untersuchungen über die Änderung von Farben unter dem Einfluss von Gestalten (Angleichungserscheinungen). *Zeitschrift für Psychologie*, 92, 249–263.
- Glaserfeld, E. von (1984). An introduction to radical constructivism. In P. Watzlawicz (Ed.), *The Invented Reality*. New York: Norton.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Hartmann, N. (1935). *Grundlegung der Ontologie*. Berlin: De Gruyter.
- Hartmann, N. (1950). *Asthetik*. Berlin: De Gruyter.
- Hartmann, N. (1952). *New Ways of Ontology*. Westport Connecticut: Greenwood Press.
- Hatfield, G. (2003). Representation and constraints: The inverse problem and the structure of visual space. *Acta Psychologica*, 114, 355–378.
- Hecht, H., & R. Mausfeld (Eds.). (2003). *Looking into pictures*. Cambridge, MA: MIT Press.
- Henry, D. (1991). *Medieval Mereology*. Grüner: Amsterdam.
- Hess, C. V. (1904). Untersuchungen über den Erregungsvorgang in Sehorgan bei Kurz- und bei langer dauernder Reizung. *Pflügers Archiv für Gesamte Physiologie*, 101, 226–262.
- Hildebrand, A. (1945). *The Problem of Form in Painting and Sculpture*. New York: G.E. Stechert.
- Holmes, N. P., Calvert, G. A., & Spence, C. (submitted). Tool-use and multisensory peripersonal space.
- Holmes, N. P., & Spence, C. (2004). The body schema and the multisensory representation(s) of peripersonal space. *Cognitive Processes*.
- Hornbostel, E. M. von (1925). Die Einheit der Sinne. *Melos, Zeitschrift für Musik*, 4, 290–297.
- Hubbard, T. (1999). How consequences of physical principles influence mental representation: The environmental invariant hypothesis. In P. R. Killeen, & W. R. Uttal (Eds.), *Fechner Day 99: The End of 20th Century Psychophysics. Proceedings of the 15th Annual Meeting of the International Society for Psychophysics* (pp. 274–279). Tempe, AZ, USA: The International Society for Psychophysics.
- Hubbard, T. (2004). The perception of causality: Insights from Michotte's launching effect, naïve impetus theory, and representational momentum. In A. M. Oliveira, M. P. Teixeira, G. F. Borges, & M. J. Ferro (Eds.), *Fechner Day 2004* (pp. 116–121). Coimbra, Portugal: International Society for Psychophysics.
- Hubbard, T., & Bharucha, J. J. (1988). Judged displacement in apparent vertical and horizontal motion. *Perception and Psychophysics*, 44 (3), 211–221.
- Hubbard, T., & Motes, M. A. (2002). Does representational momentum reflect a distortion of the length or the endpoint of a trajectory? *Cognition*, 82, B89–B99.
- Husserl, E. (1891). *Philosophie der Arithmetik: Psychologische und logische Untersuchungen*. Halle: Niemeyer. Rep. 1970, Husserliana XII, The Hague: Nijhoff.
- Husserl (1966). *Zur Phänomenologie des inneren Zeitbewusstseins*. Ed. by R. Boehm, Husserliana X. Den Haag: Nijhoff.
- Husserl, E. (1970). *Logical Investigations*. Ed. by J. N. Findlay. London: Routledge & Kegan Paul.
- Husserl, E. (1989). *Ideas Pertaining to a Pure Phenomenology and Phenomenological Philosophy*. Dordrecht: Kluwer.
- Husserl, E. (1994). Critical discussion of K. Twardowski, Zur Lehre vom Inhalt und Gegenstand der Vorstellungen. Eine psychologische Untersuchung. In D. Willard (Ed.), *Early Writings in the Philosophy of Logic and Mathematics* (pp. 388–395). Dordrecht: Kluwer.

- Ingarden, R. (1937). *O poznawaniu dzieła literackiego* (The knowledge of literary work). Lwów: Ossolineum.
- Ingarden, R. (1962). *Untersuchungen zur Ontologie der Kunst: Musikwerk, Bild, Architecture, Film*. Tübingen: Niemeyer.
- Ingarden, R. (1968). *Vom Erkennen des literarischen Kunstwerks*. Tübingen: Niemeyer.
- Intraub, H. (2002). Anticipatory spatial representation of natural sciences: Momentum without movement? *Visual Cognition*, 9 (1–2), 93–119.
- Kammer, T., Lehr, L., & Kirschfeld, K. (1999). Cortical visual processing is temporally dispersed by luminance in human subjects. *Neuroscience Letters*, 263, 133–136.
- Kanizsa, G. (1980). *Grammatica del vedere*. Bologna: Il Mulino.
- Kanizsa, G. (1991). *Vedere e pensare*. Bologna: Il Mulino.
- Khurana, B., & Nijhawan, R. (1995). Extrapolation or attention shift? *Nature*, 378, 566.
- Kirschfeld, K., & Kammer, T. (1999). The Frölich effect: A consequence of the interaction of visual focal attention and metacontrast. *Vision Research*, 39, 3702–3709.
- Koenderink, J. J., & van Doorn, A. (2002). Ecological optics and the creative eye. In D. Heyer, & R. Mausfeld (Eds.), *Perception and the Physical World* (pp. 271–304). Chichester, U.K.: John Wiley & Sons.
- Koenderink, J. J., & van Doorn, A. (2003). Pictorial space. In H. Hecht, R. Schwartz, & M. Atherton (Eds.), *Looking Into Pictures* (pp. 239–299). Cambridge, MA: MIT Press.
- Koenderink, J. J., van Doorn, A., Kappers, M. L. A., & Todd, J. T. (2002). Pappus in optical space. *Perception & Psychophysics*, 64, (3), 380–391.
- Koffka, K. (1930). Some problems of space perception. In C. Murchison (Ed.), *Psychologies of 1930* (pp. 161–187). Worcester, MA.
- Koffka, K. (1935). *Principles of Gestalt Psychology*. New York: Harcourt, Brace, and World.
- Köhler, W. (1920). *Die physische Gestalten in Ruhe und im stationären Zustand*, Braunschweig.
- Köhler, W. (1947). *Gestalt Psychology. An Introduction to New Concepts in Modern Psychology*. Liveright: New York, NY.
- Kopfermann, H. (1930). Psychologische Untersuchungen über die Wirkung zweidimensionaler Darstellungen körperlicher Gebilde. *Psychologische Forschung*, 13, 293–264.
- Knill, D. C., & Richards, W. (Eds.). (1996). *Perception as Bayesian Inference*. Cambridge: Cambridge University Press.
- Krekelberg, B., & Lappe, M. (1999). Temporal recruitment along the trajectory of moving objects and the perception of position. *Vision Research*, 39, 2669–2679.
- Kubovy, M., Epstein, W., & Gepshtein, S. (2002). Foundations of visual perception. In A. F. Healy & R. W. Proctor (Eds.), *Experimental Psychology*, Vol. 4. In I. B. Weiner (Ed.), *Handbook of Psychology* (pp. 87–119). New York: John Wiley & Sons.
- Kurana, B., Watanabe, K., & Nijhawan, R. (2000). The role of attention in motion extrapolation: Are moving objects ‘corrected’ or flashed objects attentionally delayed? *Perception*, 29, 675–692.
- Làvadas, E. (2002). Functional and dynamic properties of visual peripersonal space. *Trends in Cognitive Sciences*, 6 (1), 17–22.
- Làvadas, E., & Pavani, F. (1998). Neuropsychological evidence of functional integration visual, auditory and proprioceptive spatial maps. *Neuroreport*, 6, 1195–1200.
- Leyton, M. (1992). *Symmetry, Causality, Mind*. Cambridge, MA: MIT Press.
- Lewald, J., & Ehrenstein, W. H. (1996). The effect of eye position on auditory lateralization. *Experimental Brain Research*, 108, 473–485.
- Lipps, Th. (1879). *Raumaesthetik und geometrisch-optische Täuschungen*. Leipzig: Barth.
- Mack, A., & Rock, I. (1998). *Inattention blindness*. Cambridge, MA: MIT Press.

- MacKay, D. M. (1958). Perceptual stability of a stroboscopically lit visual field containing self-luminous objects. *Nature*, 181, 507–508.
- Marconi, D. (2005). Contro la mente estesa. *Sistemi Intelligenti*, 17 (3), 389–398.
- Marr, D. (1982). *Vision*. San Francisco: Freeman Press.
- Massironi, M. (1998). *Fenomenologia della percezione visiva*. Bologna: Il Mulino.
- Maturana H. R., & Varela, F. J. (1987). *The Tree of Knowledge: The Biological Roots of Human Understanding*. Boston: New Science Library.
- Mausfeld, R. (2002). The physicalistic trap in perception theory. In D. Heyer, & R. Mausfeld (Eds.), *Perception and the Physical World* (pp. 75–112). Chichester, U.K.: John Wiley & Sons, Chichester.
- Mausfeld, R. (2003). Conjoint representations and the mental capacity for multiple simultaneous perspectives. In H. Hecht, & R. Mausfeld (Eds.), *Looking into Pictures* (pp. 17–60). Cambridge, MA: MIT Press.
- Meinong, A. (1899). Über Gegenstände höherer Ordnung und deren Verhältniss zu inneren Wahrnehmung. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 21, 182–272.
- Meinong, A. (1960). The theory of objects. In R. M. Chisholm (Ed.), *Realism and the Background of Phenomenology* (pp. 76–117). Glencoe: The Free Press.
- Michotte, A. (1946). *La perception de la causalité*. Louvain: Etudes de Psychologie, vol. 8; En. Tr. 1963. London: Methuen.
- Michotte, A. (1950a). A propos de la permanence phénoménale: Faits et théories. *Acta Psychologica*, 7, 298–232.
- Michotte, A. (1950b). The emotions regarded as functional connections. In M. L. Reymert (Ed.), *Feelings and emotions* (pp. 128–144). New York: Mc Graw Hill.
- Michotte, A. et al. (1962). Causalité, permanence et réalité phénoménales. *Studia Psychologica*. Louvain: Publications Universitaires.
- Murakami, I. (2001). The flash-lag effect as a spatio-temporal correlation structure. *Journal of Vision*, 1, 126–136.
- Musatti, C. L. (1953). Luce e colore nei fenomeni del contrasto simultaneo, della costanza e dell'euguagliamento. *Archivio di Psicologia, Neurologia e Psichiatria*, 5, 544–577.
- Musatti, C. L. (1964). *Condizioni dell'esperienza e fondazione della psicologia*. Florence: Editrice Universitaria.
- Müsseler, J., & Aschersleben, G. (1998). Localizing the first position of a moving stimulus: The Frölich effect and an attention-shifting explanation. *Perception and Psychophysics*, 60, 683–695.
- Nijhawan, R. (1994). Motion extrapolation in catching. *Nature*, 370, 256–257.
- Nijhawan, R. (1997). Visual decomposition of colour through motion extrapolation. *Nature*, 386, 66–69.
- O'Regan, K., & Noë, A. (2001). A sensory motor account of vision and visual consciousness. *Behavioural and Brain Sciences*, 24, 939–1031.
- Oppenheim, P., & Putnam, H. (1958). The unity of science as a working hypothesis. In H. Feigl et al. (Eds.), *Minnesota Studies in Philosophy of Science*, II (pp. 3–36).
- Palmer, S. (1999). *Vision Science*. Cambridge, MA: MIT Press.
- Penrose, R. (1994). *Shadows of the Mind*. Oxford: Oxford University Press.
- Poggio, T. (1990). Vision: The other face of AI. In K. A. Mohyeldin Said et al. (Eds.), *Modelling the Mind* (pp. 139–154). Oxford: Clarendon Press.
- Poli, R. (1998). Levels. In R. Poli, & G. Scaramuzza (Eds.), *Phenomenological Aesthetics* (pp. 197–221), special issue of *Axiomathes* 9, 1–2.

- Poli, R. (2001). The basic problem of the theory of levels of reality. *Axiomathes*, 12 (3–4), 261–283.
- Rensink, R. A. (2002). ‘Change detection’. *Annual Review Psychology*, 53, 245–277.
- Rensink, R. A., O’Reagan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in the scene. *Psychological Sciences*, 8, 368–373.
- Rich, A. N., & Mattingley, J. B. (2002). Anomalous perception in synesthesia: A cognitive neuroscience perspective. *Nature Reviews, Neuroscience*, 43–52.
- Rock, I. (1983). *The Logic of Perception*. Cambridge, MA: MIT Press.
- Runeson, S. (1974). Constant velocity: Not perceived as such. *Psychological Research*, 37, 3–23.
- Shepard, R. (1994). Perceptual cognitive universals as reflections of the world. *Psychonomic Bulletin and Review*, 1, 2–28.
- Shepard, R. (1995). Mental Universals: Towards a twenty-first century science of the mind. In R. L. Solso, & D. W. Massaro (Eds.), *The Science of the Mind: 2001 and Beyond* (pp. 50–62). Oxford: Oxford University Press.
- Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neurosciences*, 18, 555–86.
- Spence, C., & Driver J. (Eds.). (2004). *Crossmodal Space and Crossmodal Attention*. New York: Oxford University Press.
- Todd, J. T., Oomes, A. H. J., Koenderink, J. J., & Kappers, A. M. L. (2001). On the active structure of perceptual space. *Psychological Science*, 12, 191–196.
- Treisman, A. (1996). The binding problem. *Current Opinions in Neurobiology*, 6, 171–178.
- Twardowski, K. (1977). *On Content and Object of Presentations. A Psychological Investigation*. The Hague: Nijhoff.
- Varela, F. J. (1999). A science of consciousness as if experience mattered. In J. Petitot, F. J. Varela, B. Pachoud, & J.-M. Roy (Eds.), *Naturalizing Phenomenology* (pp. 31–43). Stanford: Stanford University Press.
- Verfaillie, K., & d’Ydewalle, G. (1991). Representational momentum and event course anticipation in the perception of implied periodical motions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 302–313.
- Vishwanath, D. (2005). The epistemological status of vision and its implication for design. *Axiomathes*, 15 (3), 339–486.
- Watzlawicz, P. (1984). *The Invented Reality*. New York: Norton.
- Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. *Psychologische Forschung*, 4, 301–350.
- Whitney, D. (2002). The influence of visual motion on perceived position, *Trends in Cognitive Science*, 6 (5), 211–215.
- Whitney, D., & Cavanagh, P. (2000). Motion distorts visual space: Shifting the perceived position of remote stationary objects. *Nature Neuroscience*, 3 (9), 954–658.

Coplanar reflectance change and the ontology of surface perception

Dhanraj Vishwanath

Introduction

A primary goal of our visual system is thought to be the recovery of the geometry, texture, coloration, and reflectance properties of surfaces in the external environment (Gibson 1950; Marr 1982; Binford 1981; Barrow & Tanenbaum 1981; Nakayama, He, & Shimojo 1995; Anderson 2005). But what is a surface? What kind of entity is a perceived surface as a category of our visual sensorium? What kind of ontological entity is a surface as an objective physical thing in the external world? Are these two questions the same? In other words, are the properties and descriptive modes applicable to a surface specified in our ontology of the external world the same as those that our internal perceptual ontology specifies?

In the realm of the physical sciences, a surface is typically thought of as demarcating a phase-change or material interface: the boundary between two phases of matter (e.g. liquid and solid), or the boundary between two dissimilar materials in the same phase (e.g. oil and water). It is at these phase-change or material interfaces that light is differentially reflected, transmitted, and/or refracted. It seems natural that any organism that uses light as a medium to determine characteristics of its environment would exploit this interaction between light and matter. Thus, one answer to first question posed above is that a perceived surface is nothing more than a representation of the ontological entity that a real surface in the physical world constitutes. This view is consistent with what may be called the *perception-as-inference* model, where perception is viewed as the process by which entities and properties that exist in the external world are measured, detected, and encoded by our perceptual system. Under the inference model, the metrics and attributes of the perceptual encoding of a surface are assumed to be more or less the same as those that *objectively* describe the real external surface – its bounding shape, size, distance, curvature, texture, reflectance properties and coloration. In other words, surfaces in our percepts and surfaces in the world pick out entities in

the same *ontological* domain; one defined by objective physical entities, properties, and descriptors independent of perception. Marr, in his seminal book on visual perception (Marr 1980) states that the process of visual reconstruction from the retinal image is:

...the recovery of gradually more objective, physical properties about an objects shape. The main stepping stone toward this goal is describing the geometry of the visible surfaces, since the information encoded in images, for example by stereopsis, shading, texture, contours, visual motion, is due to a shape's local surface properties.

He goes on to say that

...the visible [part of the] world can be regarded as being composed of [locally] smooth surfaces, having reflectance functions whose spatial structure may be elaborate...

What is Marr saying here? He is asserting that a surface is an entity accessible to perception, which can be objectively described by its geometry and pattern of reflectance. By geometric properties, he means things like bounding shape, curvature, and relative orientation. By pattern of reflectance he means that surfaces may have an elaborate pattern of surface markings or coloration such that the light reflected is not uniform across such surfaces (for example, printed fabric, a photograph or painting, or the text on this paper that you are reading). The process of perception is to faithfully infer the objective properties of surfaces in the external world via measurement of image properties (or cues) such as image brightness, binocular disparity, gradient of texture, and the gradient of shading.¹

Problems with the definition of surfaces in inference-based models

Stroll (1982) has explained quite convincingly why a surface is more complex a notion than merely a description of a phase-change interface with objectively definable geometric structure. For example, he points out that surfaces are not perceived as impartial, abstract, interfaces between two materials. Instead, a surface appears to inhere consistently to the material that is perceived to be the *object*, suggesting a basic perceptual subjectivity in the assignment of surface ownership.² Such observations seem to go against the objectivity implied in standard inferential models of perception.

More generally, there are three drawbacks with inferential models of perception such as Marr's: they entail problematic epistemological consequences;³ they lack explanatory power in understanding the information content of the percept (Leyton, 1992); and they do not account for conspicuous phenomenological aspects of perception.⁴

Surface perception is a particularly good demonstration of these shortcomings: First, in many common situations, the very interpretation of surface layout and properties seems in conflict with any plausible inference based on the ontology suggested by Marr above. The most obvious example is the perception of pictorial space. When we look at a picture or line drawing of 3D objects, our percept vividly indicates 3D surface structure contrary to the overwhelming and unambiguous evidence for a flat, patterned, 2D surface. Such evidence is provided by binocular disparity, motion parallax, gradient of surface microtexture, and focus cues.⁵ Second, in inferential models, there is a substantive explanatory gap regarding the phenomenology of surface perception. For example, surface descriptions under such models typically reduce the perception of surfaces to the recovery of the surface's local orientations without explicit representation of other overwhelming phenomenological attributes such as the perceived *continuity* of a surface.⁶ There are also subtler aspects of phenomenology, difficult to explain under inferential theories, which are rarely broached in scientific treatments of surface perception. An important example is the perceptual effect of surface articulation that arises in the design of objects. Such perceptual effects seem to form the basis for certain generic visual preferences that are often made by designers; preferences that have been implicitly or explicitly codified into design rules (e.g., the classical rules of architectural design).

An alternative definition of visual surfaces

What sort of epistemological and ontological commitment would better capture the complex phenomena of surface perception? One view is that our perception of surfaces does not reflect an objective ontology of the external world. In other words, rather than being a *re-presentation* of an entity specifiable in an objective ontology of the external world, a visual surface is a unique ontological structure internal to our perceptual machinery. Therefore, in order to understand the nature of a surface as a perceptual entity, we need to carefully differentiate between the ontological units intrinsic to our perceptual schema or sensorium, and those entities and properties that we reflexively project onto our cognitive ontology of an external world; a cognitive ontology that is based on the very perceptual units that we are trying to understand. Making such a distinction might reveal some fundamental flaws in the ontological and epistemological claims implicit in Marr's passage; and our common folk psychology. In this paper, I propose that one such flaw is the assumption that surfaces – as defined in the ontology of our perceptual presentation – have a property that Marr called *reflectance patterns*; i.e. differential patterns of reflectance due to markings or discontinuous coloration.

According to the ontology suggested in the Marr passage, real surfaces in the external world (we will henceforth call these hypothesized external surfaces,

r-surfaces) have a property called reflectance, or reflectance patterns. The correlate of reflectance, in perceptual space, is what is called surface lightness or color.⁷ Thus, Marr's model proposes that perceived surfaces (henceforth, *p-surfaces*), have as properties, a geometry and a particular pattern of lightness/color. Under his model, the geometry and pattern of lightness of *p-surfaces* are usually (more or less) correct *re-presentations* of the geometry and pattern of reflectance of external *r-surfaces* they derive from.

What I will propose here is that *p-surfaces*, as specified in our perceptual ontology, do not have a property we might call *lightness patterns*.⁸ Instead, every perceived change in surface lightness or color, at whatever scale visible, is the presentation of a *distinct surface* (i.e. a spatially independent and identifiable surface); even when the perceived lightness change derives from a reflectance change on a single external *r-surface*.⁹

Under this proposal the perceptually presented surfaces, namely *p-surfaces*, do not have a property that we might call *coplanar lightness change*, even though we may cognitively ascribe a property *coplanar reflectance change* to *r-surfaces* in the external world; reflectance change due to differential pigmentation, patterns, or markings. In other words, a crucial property that we ascribe to surfaces in our cognitive ontology of the external world does not exist within the presentational schema of perception.

Figure 1 diagrammatically illustrates the basic claim. Imagine you are in a dark room, looking at an odd shaped planar *r-surface*, with a reflectance pattern of smaller darker colored polygons (panel *a*). The percept, under the sort of inferential model (and ontology) that Marr suggests above, is a single *p-surface* with the property of discontinuous lightness patterns as sketched in panel *b*. And this *p-surface* is an inference onto a real *r-surface* with a discontinuous reflectance pattern. The alternative ontology proposed here claims that our percept is the *presentation* of five distinct *p-surfaces* as indicated in panel *c*, four with the same lightness and one with different lightness.¹⁰ In other words, under this ontology, perception parses the sensory input of discontinuity in luminance into a presentation of geometrically distinct *p-surfaces* with uniform lightness or color.¹¹ Another way to put it would be that the perceptual encoding does not have an *ontological type* that one might call "paint".

The present claim may seem moot on one hand and contentious on the other. Moot, because the assumption of a world consisting of surfaces with homogeneous reflectance has been the basis of most successful psychophysical and computational analyses of biological vision; contentious, because on the face of it, it appears we have a strong phenomenological sense of discontinuously pigmented *r-surfaces*. I will contend that the claim is not moot because the real world does *not* consist mostly of surfaces with uniform reflectance. I will argue that the claim is not contentious if one is able to differentiate between our immediate percep-

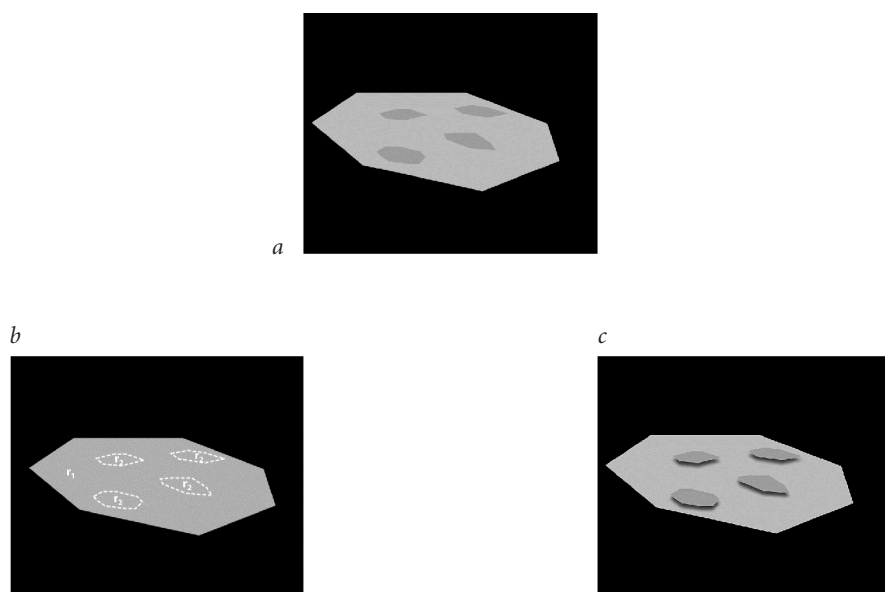


Figure 1. Perceptual interpretations of a surface with coplanar reflectance change. (a) an image of a simple surface-object with coplanar reflectance change. (b) and (c) are two possible interpretations. The interpretation (b) is a single planar surface with a secondary property of a discontinuous reflectance pattern indicated by the dashed regions (r_1 and r_2 indicate the two reflectance values). In (c) the object is interpreted to be made up of five distinct, nearly coplanar, surfaces of uniform lightness (the shadows and exaggerated separation are for illustrative purposes only).

tual phenomenology and the secondary cognitive interpretations we then apply in categorizing our world.

Under this proposal, all change in perceived brightness¹² is the internal presentation of either continuous change of pose (e.g. a curved surface) or change of identity of p-surfaces. The presentation of identity change may either take the form of a percept of discrete p-surfaces with different lightness (Figure 1c), or the presentation of discontinuous change of pose (or spatial position) between two p-surfaces of the same perceived lightness (left and right vertical surfaces in Figure 2).

P-surface lightness under this account is not a re-presentation of surface reflectance (as implied in traditional inferential accounts of perception), or just a secondary feature or property in the perceptual representation of surfaces.¹³ Perceived discontinuous lightness change is the *very presentation* of surface identity.

So what then is the perceptual phenomenology associated with surfaces that have lightness change but appear coplanar, such as we seem to have in Figure 1a? I will suggest that such a percept exemplifies a state of *representational conflict*; a

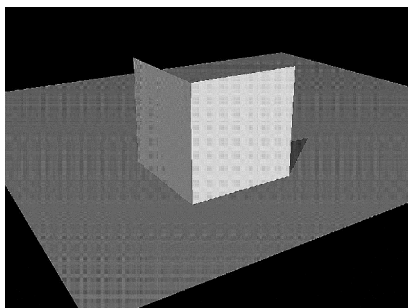


Figure 2. The image of an object made up of two surfaces with the same reflectance abutting at a right angle. The difference in brightness of the two surfaces is due to the difference in orientation of the surfaces with respect to the light source, resulting in lower illumination on the left facing surface. The perceptual interpretation is of two surfaces with similar lightness with a discontinuous change of pose at their adjoining edge.

state where there is a high degree of internal conflict in the ontological and structural (or geometric) parsing intrinsic to perceptual presentation.¹⁴ What are the consequences of such ontology and the resulting perceptual conflict? I will argue that they have critical qualitative consequences in the design of artifacts and environments that involve articulation of surfaces.

Surface ontology in inferential approaches to perception

Let us now look more closely at inferential models of surface reflectance and lightness perception. Perception-as-inference approaches assume an objective external 3-dimensional world of objects, surfaces and properties that can be described by certain observer-independent descriptors and measures. The process of perception, then, is to infer those objects, surfaces, and properties from a 2-dimensional image of the world that is captured at the sensory interface.

In the visual domain, this is often called the *inverse optics* problem because the process of inference can be thought of as reversing the *forward optics* that result in the projected retinal image (Figure 3). Since the inferred 3D entities are specified in a higher dimensional space than the image – i.e. the inverse mapping is one-to-many – the perceptual apparatus has to apply certain constraints to make the inverse-mapping computation tractable. The constraints can be thought to be assumptions about the behavior and structure of the world that are required in order to uniquely infer properties of the world. There are 3 aspects that are typically considered in the inverse-optical inference of surfaces:

Illumination. This is the pattern of light impinging on the viewed r-surface from various light sources present.

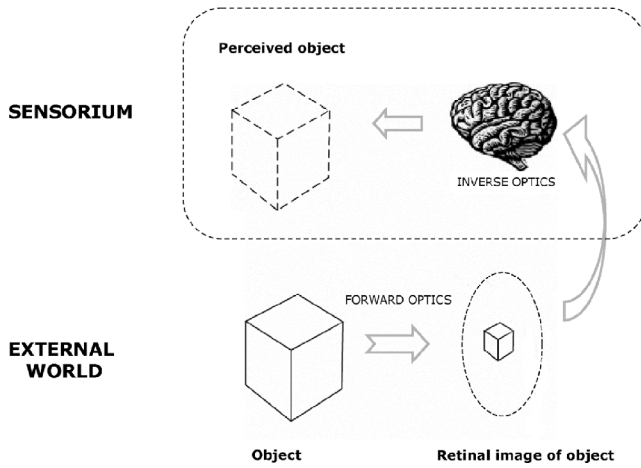


Figure 3. Perception as inference: A simple diagram depicting the inverse-optics model of perception.

Surface reflectance. Reflectance is the reflective property of the r-surface in the external world that is to be inferred. This property arises from the pattern of pigmentation and micro surface structure. The perceptual correlate of reflectance is called *lightness*.

Surface luminance. This is the overall amount of light emanating from each part of the external r-surface toward the eye. The actual point-wise intensity of light arriving at different parts of the retina will naturally be directly proportional to the r-surface luminance. Surface luminance will depend on (1) r-surface geometry, i.e. the relative orientation of the surface region with respect to viewer (2) the reflectance properties of the r-surface (3) The pattern of illumination. The inferred or perceived p-surface luminance is known as *brightness*. (For simplicity I will use the term brightness interchangeably to also describe the intensity of the retinal image.)

Surface geometry. Geometry is usually represented in terms of a surface's bounding contour shape, and the slant and tilt of the surface with respect to the observer.¹⁵ For a planar surface, slant may be expressed simply by the orientation of the central surface normal with respect to the observer. For an extended non-planar surface, slant may be expressed in terms of curvature, or slant and tilt may be enumerated separately for each infinitesimal patch making up the surface. The overall surface geometry is usually called *surface shape*. In inferential models of perception it is assumed that p-surface shape will be (more or less) the same as r-surface shape, assuming the inference is correct.

The recovery of the true properties of the r-surface involves the inference of surface geometry (surface shape) and its reflectance properties from the measured

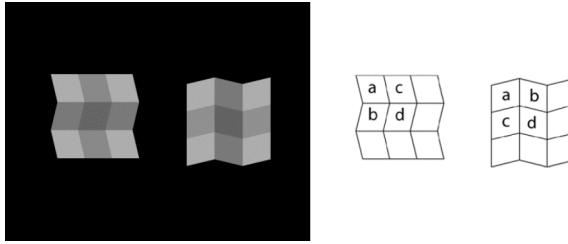


Figure 4. The left panel is an image depicting two surface-objects. The objects appear to be light-colored bent rectangular surfaces with a dark stripe down (or across) the middle. The right panel identifies the different regions of the image (see text). Adapted from Adelson & Pentland (1996).

intensity of light at different parts of the retinal image. In order to do this, the process must correctly discount the real illumination. In other words, any luminance variation due to the nature of the illuminant must be discounted such that only effects due to the geometry and reflective properties are used in recovering surface shape and reflectance.

Consider you are in a completely dark room looking at the two objects depicted in Figure 4 (left panel); two bent, light-colored surfaces with a dark stripe running down the middle. The luminance of patches a, b, c and d on the actual object will be different, though the reflectance of patches a and b (or c and d) are the same. Similarly, the *perceived* brightness of the surface patches a, b, c and d are all different but the patches a and b (or c and d) are perceived to have the same lightness; we infer the difference in brightness between a and b (or c and d) to be due to the change in spatial orientation between the respective patches. A standard interpretation of our percept of these two objects is that in each case we perceptually infer a single light-colored rectangular surface with a dark stripe painted down the middle that has been bent at two places, as illustrated in Figure 5 (where the dashed lines indicate the location of the bends).

How might the visual system go about making such an inference of reflectance change and geometry from the actual pattern of luminance that is sensed by the eye? Naturally, the inference must select a set of values for geometry and reflectance change consistent with the various cues available to the visual system. The cues to surface reflectance are the measured pattern of brightness at the retina.¹⁶ The cues to surface structure are measured binocular disparity, motion parallax, texture gradients, gradients in brightness, etc. In order for the visual system to infer an optimal or veridical percept it needs to integrate these measurements so that the final inferred structure is consistent with the individual cues.¹⁷

One way to describe such a process is within a Bayesian probabilistic model, where the goal is to determine the probability of observing a particular environ-

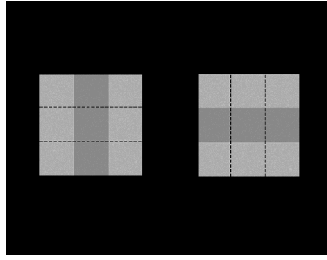


Figure 5. The standard perception-as-inference interpretation of the objects in Figure 4. Each object is a light colored planar surface bent at two locations (along the dotted lines) with a stripe of darker color painted down (or across) the middle.

mental configuration of surfaces given the various sensory measurements, or cues, available in the image. This probability, called the posterior, is derived by determining the probability that a particular surface configuration could have given rise to the measurements in the image (the likelihood) and the intrinsic probability of observing such types of surface configurations (the prior). The correct interpretation is the surface configuration that has the highest combined likelihood and prior probability, i.e. the highest posterior probability. This approach is akin to a hypothesis-testing model, where the hypothesized solution that best explains the image information is inferred to be the correct one.

Adelson and Pentland's workshop metaphor

Adelson and Pentland (1996) have provided a particularly good characterization of the Bayesian approach to surface inference in what they call the “workshop metaphor” for surface perception. Their example is relevant here, not only because it is computationally well motivated, but because it also suggests a simple ontological conceptualization of the problem. In their model, the problem of parsing observed image intensities (brightness values) into an interpretation of geometry (surface shape) and reflectance pattern (perceived lightness pattern) can be thought of in terms of a set of “specialists” who each contribute to the interpretation of *part* of the retinal image information, and can collaborate to find the best 3D inference of surfaces given the entire image. Here is the description provided by Adelson and Pentland:

We begin by describing a “workshop” metaphor. Suppose that we are given the task of constructing a physical scene that will produce the image of [Figure 4]. We go to a workshop where a set of specialists build the scenery for the stage sets used in dramatic productions. One is a lighting designer; another is a painter; and a third is a sheet-metal worker. There is also a supervisor who can coordinate the actions of the individual specialists. We show them the desired image, and ask them

to determine how to build a scene that will look the same. They are faced with a problem analogous to the one faced by the human visual system: given an image, try to figure out how it could have come about. Let us imagine that the specialists charge according to a set of fixed prices. Simple and common operations are cheap, while more complex and unusual operations are more expensive. We can then cast the [workshop's] problem in terms of minimizing a cost function. The notion that a percept should correspond to the simplest or likeliest explanation of a scene has a long history in the perception literature. . . and it has more recently been shown that formal concepts of simplicity (e.g. minimal length descriptions) and likelihood (e.g. maximum likelihood estimators) are fundamentally related. . . These approaches can both be formalized as minimizing a cost function.

We saw earlier how inferred surface properties depend on three properties of the scene: illumination, reflectance, and surface geometry, and an optimal integration of the cues available in the image with respect to each property. These are translated in Adelson and Pentland's model respectively into a lighting designer, a painter, a sheet-metal worker, and a supervisor. Each expert charges a certain amount for any task he/she performs as shown in Table 1. The cost of the final configuration will depend on what each specialist contributes in constructing the final configuration that can explain the image. Naturally, there are many different ways of constructing scenes that could produce the image. One scenario is where each expert constructs the scene almost entirely by himself. For example, the painter could create a scene consistent with the image by simply painting the appropriate shapes with different shades of paint on a flat sheet of metal and have lighting designer illuminate it with a single flood as illustrated in Figure 6a. Similarly the sheet metal worker could create a scene that just uses sheet metal; cut, bent, and oriented with respect to a single flood such that the brightness differences are achieved by the differences in shading due to differences in orientation of the metal panels (Figure 6b). The lighting expert could just set up 9 custom spotlights that cast lights of the appropriate shapes and varying intensity onto a white screen (Figure 6c). A "cost accounting" of each of those solutions is shown below in Table 1.

The perceptual analogy for each of these solutions would be an interpretation of the image in Figure 4 consistent with (1) a single flat surface with patches of different lightness; (2) a complex spatial configuration of surfaces of uniform lightness; or (3) a single flat surface with uniform lightness appeared to be illuminated by light sources of different intensities. However, as mentioned earlier, when we view the corresponding object in the left panel in Figure 4, what we seem to perceive is a single bent surface with a dark stripe down the middle. Adelson and Pentland explain how such a percept can be explained in the workshop metaphor as a collaborative low-cost solution determined by a supervisor utilizing the skills of each specialist and weighing their relative costs. That solution and its cost are

Table 1. “Costs” related to various solutions to the generations of the image in Figure 4, produced by the metaphorical workshop. The upper panel shows costs for each specialist working separately. The lower panel shows the cost of a supervised solution involving all the specialists (see text below). From Adelson & Pentland (1996).

Painter’s solution	Sheet metal worker’s solution	Lighting designer’s solution
Cut 1 rectangular sheet = \$8 Paint 9 polygons = \$180 Set up 1 flood = \$5 Total = \$193	Bend 6 odd angles = \$30 Cut 24 sides = \$120 Set up 1 flood = \$5 Total = \$155	Cut 1 rectangular sheet = \$8 9 custom spots= \$270 Total = \$278
Supervisor’s solution		
Cut 1 rectangle = \$8 Paint 3 panels = \$60 Make 2 bends = \$10 Set up 1 flood = \$5 Total = \$83		

shown in Table 1 as the *supervisor’s solution*. In a similar vein, we can think of the perceptual interpretation of the objects in Figure 4, shown in Figure 5, as an optimal, low-cost, solution for the perceptual system.

The actual implementation of Adelson and Pentland’s model involves an iterative mixture-of-specialists algorithm that searches for the best solution given certain predetermined constraints placed on each specialist function, in what they call intrinsic image space.

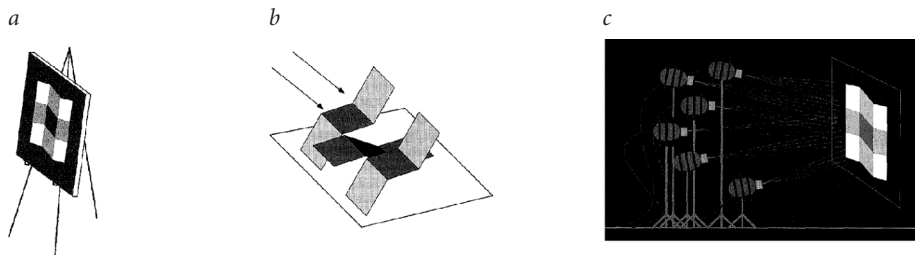


Figure 6. Three potential solutions to producing the image in Figure 4, as proposed by each of the specialists in the perceptual “workshop” working separately. From Adelson & Pentland (1996).

Problems with inferential approaches to surface perception

Though the kind of model presented by Adelson and Pentland has quite useful application in the arena of machine vision, when we attempt to stretch the analogy to human perception, several problems arise. In the workshop metaphor, the particular costs for the specialists' tasks – and correspondingly, the cost functions applied in the computational model – were chosen so that the lowest-cost outcome would match what we perceive when we view the image in Figure 4. Now consider the situation of viewing “real” 3D surfaces in a dark room (rectangular sheet with two bends and a darker stripe painted down the middle as depicted in Figure 4) – the Adelson and Pentland's interpretation of cost functions seems at first blush to be reasonable. The perceptual system might be thought to have the following bias: Explain as much of the intensity variation at the retina as changes in surface geometry; whatever can't be explained by surface geometry should be explained as reflectance change; and (for more complex images) any information that can't be explained by either of those two, should be explained as change in illumination. This bias seems to explain why we might arrive at a correct percept when we view a set of real objects as depicted in Figure 4. But now consider the left panel of Figure 4 as an image on the piece of paper that you are reading. Its “real” physical status is a flat surface with patches of different reflectance, yet we vividly perceive a 3D interpretation of surfaces, where one edge appears at a different depth than another edge. In other words, we perceive what we would perceive if we were viewing the actual 3D objects depicted. Is this consistent with an Adelson and Pentland type cost model?

On one hand it is. That's because Adelson and Pentland restricted their model to arrive at an inference that “explains” only the information provided by the *brightness variation* in the image. No other sources of information are accounted for in their model. However, when we actually view Figure 2 with both eyes, there is a whole range of information in addition to the brightness variation that also needs to be “explained” by the workshop. For example, the disparity information specified by the difference in the two eye's images, any parallax information due to head motion, the visible texture of the paper (or computer monitor), the convergence of the eyes, and the accommodative state of the lens. Any inference model such as that suggested by Adelson and Pentland should account for all these sources of information. In other words, any reasonable cost-based model should not only deliver an inference that is “low cost” or minimal, but also one that is consistent with the available information within the inverse-optics framework in which it operates. Yet *all* these sources of depth information are entirely inconsistent with the 3D interpretation we see when we look at Figure 4, since all these sources of information signal the planar surface of the page. The percept – a non-planar surface – is therefore completely contradictory! In other words, under a cost-based

model, the 3D interpretation that we perceive – and the one that Adelson and Pentland’s model arrives at – is in a probabilistic sense, a very *high cost* interpretation. An interpretation that is inconsistent with all independent sources of information regarding surface geometry is forced, despite the fact that an interpretation that is consistent with that information exists with an Adelson & Pentland-type (or Marr-type) model – namely a percept of a single flat surface with a discontinuous lightness pattern. The problematic nature of the automatic inference of 3D structure from a physical 2D image is not limited to Adelson and Pentland’s model, but any inverse-optics based computational model of perception that takes the percept to be a minimal and optimal interpretation of the image information.¹⁸

What might be the reason for the non-optimal percept we have when we view Figure 4? One explanation might say that coplanar reflectance change is a special case: a cue-conflict stimulus. The argument may go thus: such stimuli are inherently ambiguous because the so-called monocular cues to depth in the image (brightness variation, texture gradients, and perspective) are conflicting with other cues to depth such as binocular disparity, motion parallax, accommodation, etc.; and therefore such a stimulus constitutes a degenerate case where we would expect the interpretation to be non-optimal. Let us examine this explanation.

Is coplanar reflectance change a cue-conflict stimulus?

Cue conflict is a term used to describe visual stimulation where the estimates of depth from different sensory measurements don’t agree. For example, in a stereo image shown on a flat display screen (e.g. using stereo goggles) the accommodative (focus) cue to distance is in conflict with the vergence and disparity cues to distance. Cue-conflict situations can also be deliberately created in order to understand how various cues to depth might be calibrated in the 3D percept, for example by pitting motion parallax information against disparity information (Allison & Howard 2000). Under such conditions, the visual system arrives at an interpretation where the metrics of the percept is usually some compromise between the signals to depth from the conflicting cues. But though cue conflict can be created in the laboratory or in simulated environments, it is virtually non-existent in the real world. It is certainly not the case that surfaces with coplanar reflectance change present any sort of geometric cue conflict, particularly under the model of surface ontology and inverse-optics inference that Marr, or Adelson and Pentland, specify. For such inferential models of 3D perception there is a valid interpretation that is *entirely consistent* with the various sources of information for geometric structure and surface brightness. For the left panel of Figure 4, which is a flat 2D image, that solution would be the one proposed by the “painter” in Adelson and Pentland model; i.e. an interpretation of a flat surface with discontinuous coplanar reflectance change. In the Adelson and Pentland workshop model, the interpre-

tation of a surface with two bends and a stripe painted down the middle seems reasonable because they have presented a model that does not take into account geometric information that is available when the 2D image is viewed binocularly by a human observer (i.e., disparity, convergence, motion parallax, and gradient of focus). Yet the core principle of the Adelson & Pentland model is that the interpretation must explain all variation in the image(s), and should do so in the simplest way or lowest cost possible. Therefore, given the underlying assumptions in their model, the only reasonable low-cost solution, once we add in the additional depth information, would be the painter's solution (Figure 6). Using their workshop metaphor, we might say that the supervisor would not *allow* the sheet metal worker to make any bends, however cheap, because any bends would conflict with information in the image! In probabilistic terms we would say that the *likelihood* of the image information (disparity, parallax, etc.) provided by the left panel in Figure 4, given the 3D configuration of a single bent surface, is *zero*.

Is the interpretation of coplanar reflectance ecologically adaptive?

Another plausible explanation for the contradictory percept in Figure 4 is that the inferential device has certain built-in biases that force a preference for certain interpretations. In the present case one might claim that the visual system has developed a powerful inferential bias for surfaces with uniform reflectance. Thus, any brightness change in the image is preferentially interpreted as being due to the change in orientation of a surface, rather than change in reflectance, despite strong evidence otherwise. Once we assume such inferential bias, interpretations that imply less reflectance change, and correspondingly more geometric change will be optimal from a probabilistic viewpoint. But why would the visual system have such a bias or prior? One explanation might be that it reflects a preference or adaptive bias learned from interaction with the environment. For example, in a probabilistic model of inference, a uniform-reflectance bias may be due to ecological statistics: the prior probability of uniform reflectance surfaces in the external world might be very high; which is equivalent to claiming that the prior probability of coplanar reflectance change must be very low. That would be a great explanation except for the fact that the visual world of relevance to organisms is literally brimming with coplanar reflectance change!

The visual systems of vertebrates and invertebrates have evolved alongside a fundamental visual characteristic of flora and fauna: camouflage. Figure 8 shows three examples of coplanar reflectance change in organisms (an owl moth, a zebra, and a poison-arrow frog). Camouflage is, needless to say, the example par excellence of coplanar reflectance change in the external ecology. Indeed, some have suggested that certain visual functions, e.g. stereovision, are actually adaptations that have evolved to deal with camouflage (Julesz, 1971).



Figure 7. Camouflage: three examples of naturally occurring surfaces with coplanar reflectance change. An owl moth, a zebra, and a poison arrow frog (owl moth image © Terry Carr, courtesy of Terry Carr; zebra image © The Wellington Zoo, New Zealand, with permission; poison arrow frog image courtesy of Dr. Craig Roland).

Camouflage and the argument against ecology

What is camouflage? It is essentially an adaptation in the organism's external coloring that takes advantage of an intrinsic inability of animal visual systems to distinguish between brightness changes due to change in geometry or identity of surfaces, from that due to coplanar reflectance change. Figure 8a is a photograph of a poison-arrow frog shown against some leaves. The coplanar reflectance pattern on the surface of the frog's skin is designed such that the predator's visual system will interpret the visual stimulus as shown in Figure 8b. Here, the identity of the frog's body surface is broken into several smaller surfaces with different shapes (similar to the background surface structure of leaves). This prevents the identification of the frog whose real surface shape is shown in Figure 8c. A counter adaptation in the visual system of a predator interested in identifying the frog would be to evolve mechanisms that prevent coplanar reflectance change from being interpreted as structural change in surface shape. Such a visual system may then generate a percept that is sketched in Figure 8d, where the visual stimulus is correctly parsed into the identity of a single p-surface; and were reflectance change becomes a description of a *secondary property* of that single surface (similar to what was sketched in Figure 2b).

Yet, as we have seen with Figure 4, the most evolved visual system in the animal kingdom, the human brain, quite easily succumbs to the interpretation of coplanar reflectance change as structural or geometric change. It does so despite having the mechanisms to defeat such non-optimal percepts. Both motion parallax and binocular disparity provide excellent signals to prevent any 3D interpretation of 2D reflectance change.

This suggests that an ecological argument does not explain why we see reflectance change as structural or geometric change. Neither a prior preference explanation based on ecological statistics, nor one based on adaptive function, appears to agree with the facts. Uniform-reflectance does *not* necessarily have a high



Figure 8. (a) the image of a poison arrow frog against some leaves. (b) through (d) are possible perceptual interpretations of the image by a predator visual system. (b) coplanar reflectance change on the frogs body is interpreted as change in surface structure. Note that this results in an incorrect interpretation of surface shape; precisely what the camouflage is designed to do. (c) The coplanar reflectance change is ignored and the correct surface shape of the frog's body is perceived. (d) The interpretation suggested by a perception-as-inference model. The shape of the frog's body is correctly identified and the surface pattern is interpreted to be a secondary property of the surface.

prior probability, since the visual ecology relevant to organisms is brimming with surfaces of coplanar reflectance change.¹⁹ Any adaptive bias in a predator or prey visual system should have evolved to *defeat* the tendency to interpret 3D structure from reflectance change, not *encourage* it!

The conclusion from camouflage is that an adaptive inference model of perception should predict exactly the opposite of what the human visual system sees. The fact that our visual system infers 3D structure from coplanar reflectance change, in spite of available mechanisms such as stereopsis, indicates that there is something more fundamental in the perceptual schema that does not *support* a representation of surfaces with coplanar reflectance change. In Bayesian terms we may say that the prior for coplanar reflectance change has to be zero.

An alternative ontology (Or why Adelson & Pentland have to fire the painter from their workshop)

The ontology suggested in inferential models of perception such as Marr, or Adelson and Pentland, implies that our percept of surfaces is something like that illustrated in Figure 1b (or in the case of the frog, like Figure 8d); where surfaces have a primary property of geometric shape and a secondary property of a particular reflectance or reflectance pattern. But if such an ontology is correct, then all ecological and adaptive explanations suggest that our percepts should be biased toward interpreting coplanar reflectance change *to be* coplanar reflectance change. Instead, the visual system seems strongly biased toward interpreting coplanar reflectance change as geometric change, resulting in interpretations that are closer to those shown in Figure 1c and 9b.

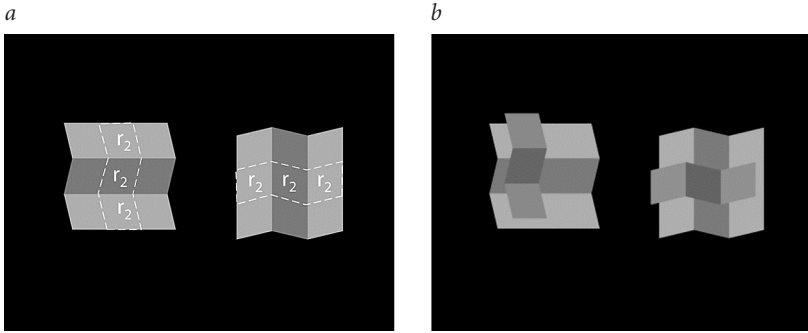


Figure 9. Two alternative ontologies in the perception of coplanar reflectance change. (a) In the perception-as-inference approach, the objects are interpreted to be single surfaces with a secondary property of reflectance patterns (r_2 being the inferred reflectance of the darker stripe). (b) In the alternative ontology, each region with a different lightness is interpreted to be a distinct surface.

I have claimed that the reason we parse an image such as Figure 1, into the interpretation (c), rather than (b) is because the ontology suggested in Figure 1b *does not exist* within the perceptual encoding. Instead, perceived change in surface brightness and lightness/color is more correctly thought of as a primary phenomenal property intrinsic to the perceptual identity of p-surfaces. Surface lightness, I am suggesting, has evolved as a phenomenal marker of p-surface identity, independent of the other two markers of identity, location and geometry.

In the standard inference model of perception, reflectance change is implied to be a property of the recovered surface. Thus when we view Figure 4, what we should see is something like what is sketched in Figure 9a, a bent flat p-surface with a secondary coplanar lightness pattern (r_2) along the middle. In the alternative view, since lightness patterns – as a *property* of p-surfaces – do not exist, all brightness variation in the image is parsed into presentation of surfaces with different structural identity and/or spatial layout (Figure 9b) even when such a parsing is inconsistent with sensory measurements.²⁰

Under this new ontology, change in identity due to change in brightness is rendered spatial or geometric wherever possible (e.g. as a change in surface orientation, due to a bend or break). However, at certain boundaries, a change in surface identity obtains without a resolvable change in spatial location. Thus, the strip in the middle of Figure 9b is perceptually *identified* as a separate p-surface, but its geometric separation from the underlying lighter p-surface, is perceptually infinitesimal. This generates a conflict between identity given by the change in surface lightness and the lack of identity in spatial position. This is what I have referred to as *representational conflict*. Note, however, that since the perceptual ontology forces a change in surface identity, the dark strip cannot correctly

be thought of as a surface pattern or marking (i.e. a secondary property of the underlying surface) as implied by our naïve externalist ontology.

This alternative ontology suggests that Adelson and Pentland would have to *fire the painter from their workshop!* In the new workshop, the sheet metal worker creates all the entities, albeit from different shades of sheet metal. He has to “fake” the stripe down the middle, by bonding the dark metal sheet to the light metal sheet as best he can.²¹ The claim here, in perceptual terms of surface identity, is that the metal worker – though he tries – cannot *in principle* bond the two surfaces such that the identity of the top dark surface gets absorbed into the identity of the underlying lighter surface. The identities remain separate at the perceptual level.

Let us summarize what we’ve discussed so far. Theories of perception-as-inference suggest that the ontology of our percepts matches an objective ontology of the external world. Adelson and Pentland’s workshop metaphor provide a useful ontological breakdown for surface perception based on our naïve-realist view of objective r-surface properties. Surfaces, in such ontology, are specified as objective entities with a particular geometry, and a potentially complex pattern of reflectance. The inferential process of perception is supposed to correctly recover entities and properties of the external world defined under such ontology, where reflectance change is inferred as lightness change. We have seen how our actual perceptual interpretations appear to go against any sort of optimal, adaptive, or ecologically valid inference based on such ontology. I introduced an alternative perceptual ontology where surfaces do not have the property of coplanar reflectance or lightness patterns. Under this ontology, surfaces are, by definition, of uniform lightness; and differences in perceived lightness are instead solely an indicator of categorical identity of p-surfaces in perception. The percept of p-surface lightness and chromaticity (the putative correlates of reflectance) is the product of a perceptual schema that provides for surface identification above and beyond the natural perceptual identity afforded by location and shape. The most important aspect of this idea is that the ontological parsing into discrete p-surfaces is *primary*; in that it will force percepts that may be in conflict with all the various measures (cues) that can be identified in the percept. Also, the presentation of the percept (e.g. Figure 9b) is to be interpreted as a conflict in *perceptual presentation* and not *measurement* (i.e. cue-conflict).²² This will become clearer in the next section.

The vivid and ubiquitous experience that we appear to have of discontinuously pigmented but geometrically continuous surfaces is, under this schema, a cognitive inference dependent on higher faculties, measurement and/or convention.²³ For example, higher cognitive functions will naturally be able to “label” or categorize surfaces as having the additional non-geometric surface property of perceived lightness/chromaticity, or pattern of lightness. But at the perceptual level, lightness is not a label, or property, attached to the geometric description of the surface; it is the very presentation of surface identity.

Since coplanar reflectance change is ubiquitous in the visual environment, we next address the *qualitative* consequences of such visual configurations and of surface perception more generally.

Coplanar reflectance change and representational conflict

Discontinuous lightness change, under the new ontological schema proposed, necessarily implies change in surface identity. Thus, the perception of coplanar lightness change implies the percept of multiple identifiable surfaces. Identity of a spatial entity requires (at a minimum) an identity in spatial position, in other words two objects cannot occupy the same spatial position or extent.²⁴ Coplanar lightness changes present a perceptual conflict, because they imply change in identity of surfaces without a corresponding spatial distinction. However, a *physical* stimulus having coplanar *reflectance* change is not sufficient for representational conflict to occur. Representational conflict requires the *perception* of coplanar *lightness* change. Figure 10 demonstrates the distinctions between coplanar lightness change and coplanar reflectance change. The panel *a* in Figure 10 can be considered to be a pattern of coplanar reflectance changes on an *r*-surface (the paper on which it is printed). The patches l_1 and l_2 in Figure 10a have different *physical* reflectance in the physical image, as do the patches d_3 and l_3 . The pattern of lightness change in the *perceived p*-surface layout however is quite different. The patches d_1 , d_2 and d_3 appear to all have the same lightness, (as do l_1 , l_2 and l_3) but the *d*-patches have a different lightness than the *l*-patches. Even though the reflectance change between l_1 and l_2 (or l_2 and l_3) is *coplanar* in the physical stimulus (the paper on which it is printed), in the perceived configuration they appear oriented differently and have the same lightness. Thus they do not constitute perceptual conflict, because in the perceptual domain the difference in brightness is resolved into a spatial property – the change in orientation between l_1 and l_2 . But there *is* a perceptual conflict due to the lightness change between the *d* surfaces and the *l* surfaces. The identity between the *d* surfaces and *l* surfaces (say d_1 and l_1) since their lightness difference is not supported by the key spatial property of identity, namely location. Therefore, while the ontological presentation is that l_1 is a distinct entity from d_1 given by their difference in lightness, the structural presentation is that they are *co-located*.²⁵ In other words, there is a conflict in the perceptual presentation.

The identity of the *d* surface patches (or *l* surface patches) in relation to the identity of the underlying image surface (the paper on which you view Figure 10a) does *not* constitute representational conflict. This is because at the level of perceptual presentation, the *d* and *l* surfaces are not *perceived* to be co-located with the actual paper surface on which Figure 10a lies. They appear to be at a different

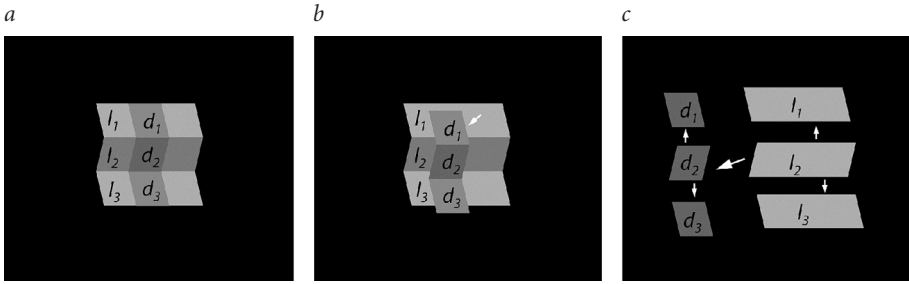


Figure 10. Perceptual presentation of an image with coplanar lightness and discontinuous brightness change (same as Figure 4). (b) The l surfaces are perceived as being of a different lightness than the d surface patches, and therefore are identified to be discrete surfaces. The differences in brightness within the d and l surface patches make each element appear to be at different orientations, but connected. (c) the current conjecture claims that even these surface elements are identified as discrete in the perceptual presentation despite their connection.

orientation and depth (most likely they appear to be farther away than the actual page). In other words, though the stimulus (the image on the paper) represents coplanar reflectance change, the d and l surfaces do not represent coplanar lightness change with respect to the paper in the perceptual domain; they have been parsed into p-surfaces with an independent geometry, location and lightness.²⁶ On the other hand, the d surfaces do present coplanar lightness change with respect to the l surfaces, and thus do present a perceptual conflict.

This is a very important observation, because it implies that 3D pictorial space is, in principle, not a condition for representational conflict, despite the fact that it is generated from a physical surface with coplanar reflectance change. Even though the measures of depth such as disparity and motion parallax are in conflict with the perceptual interpretation of pictorial scenes or objects, there is no intrinsic perceptual conflict in the ontological and structural presentation of 3D surfaces in the scene in relation to the supporting picture surface.²⁷ Of course, if there are surfaces in the picture that are perceived to have coplanar lightness change, then those surfaces will present a representational conflict within the pictorial space!

Discontinuous brightness change and representational conflict

There is an additional distinction between the interpretation that Adelson and Pentland's model provides, and the alternative ontology that I have proposed. Recall that according to Adelson and Pentland's model, the identity across the surfaces, say l_1 and l_2 , does not change because they are seen as parts of a single surface that have been bent. However, under the alternative proposal these sur-

faces are parsed as being *distinct*. In other words, the identity of surfaces across the geometric discontinuities (the bends) also changes. Unlike the description in Adelson and Pentland's model, all *discontinuous* brightness change is also rendered into a change in identity, such that (say) the lighter *l* surface is not perceptually parsed as a single bent surface, but as three different surfaces as shown in Figure 10c.²⁸ Therefore, the alternative ontology suggests that the sheet metal worker in Adelson and Pentland's workshop *cannot* make discontinuous bends, because that would deny the categorical identity of the individual surface elements (d_1 , d_2 , l_1 , etc.) that make up the overall surfaces *d* and *l*. Instead, the sheet metal worker has to try to "glue" together discrete sheets for each surface element, as best he can, in order to generate the desired surface configuration. Thus, we see that the phenomenology of perception suggest that two assumptions in Adelson & Pentland's workshop might not hold. The painter does not exist and the sheet-metal worker does not have the tools to bend sheet metal.

More generally, I will suggest here that representational conflict is intrinsic to any percept of surfaces regardless of whether the identity of surfaces is defined by a change in surface geometry, or lightness alone. So we will now consider how even surface configurations where the identity of surfaces is solely due to spatial or geometric factors can still present a representational conflict.

A continuous planar or curving surface is perceptually identified as a single surface despite the continuous brightness change. But two surfaces patches that are connected at some part of their boundary have separate identity. Think of the two walls that connect at a corner in the room you are in. It is clear that their identity as surfaces is distinct; but how distinct? In other words, at what level of *structural resolution* do we consider them distinct? I propose that connected surfaces also present an inherent conflict in perceptual presentation. When you look at the corner of the room there is a conflict between the categorical presentation of the identity of the walls as distinct entities, and the structural presentation that is unable to achieve an infinite resolution in the expression of their geometric discontinuity.²⁹ More generally, there is a *categorical* presentation of a change in *identity* of surface across any curvature discontinuity but *not* a categorical demarcation in the underlying structural presentation. This is what generates a perceptual conflict.³⁰ This conflict is independent of the source of the perceptual stimulation, pictorial or real. The resolution of the structural distinction of surfaces might be said to increase with the level of perceived discontinuity between the surfaces, so for example, in Figures 11a–c, the structural resolution would be greater in the configurations toward the right; however, the identity of two surfaces in all three remain categorically different in every case. Therefore we would predict that the representational conflict decreases as one moves from left to right.

We might push this line of thinking a little further and claim that, in general, the nature of the demarcation between surfaces is defined not only between the

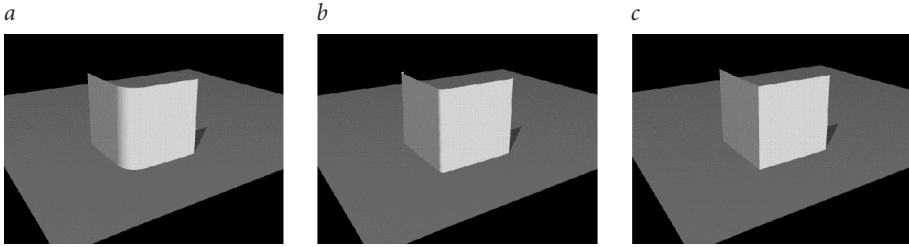


Figure 11. (a–c) images of two connected surfaces. The curvature discontinuity between the two surfaces increases from the left to right panels. In each panel the two surfaces are perceptually identified as separate, but they have differing degrees of structural identity. Under the alternative ontology of surface perception discussed in the text, the perceptual presentation of the rightmost image (d) has the lowest representational conflict.

adjoining surfaces, but also in contrast to the perceptual “ground”. Thus, we might predict that even spatially distinct surfaces have a level of representational conflict depending on how clearly they appear as structurally distinct from the perceptual ground. In every case, however, the perceptual presentation of identity remains categorical.

Representational conflict and design

In order to understand how representational conflict may have functional consequences in vision, we first consider some general aspects of the operation of the visual system in attaining a 3D visual percept. The visual systems reaction to sensory input involves a dynamic adjustment of both the optical plant as well as the 3D perceptual representation. One can think of these dynamic responses to stimulation as those that seek equilibrium states.³¹ Classic examples of adjustments of the optical plant include accommodation to sensed defocus, pupillary response to light level, and changes in the ocular vergence of eyes to match the perceived distance of fixated objects. Another part of the dynamic visual response involves the adjustment of the metrics of the 3D percept in response to depth cues (disparity, motion parallax, gradient of blur etc.), such that the metrics of the final percept are internally consistent with information presented by each cue.³² For example, it is well known that metric estimates of depth or slant from binocular disparity can take about one second to resolve.

Under normal operating conditions of the human visual system, the dynamic responses of both the optical plant as well as the generation of the perceptual product (say a perceived object, with its attendant attributes and metrics) achieve what one might call a “resolved state”. For example in the case of the accommodative

response, the dynamic adjustments end up achieving a clearly focused image at the point of fixation. In the case of perceived metrics of an object, the resolved percept usually presents an optimal interpretation consistent with the measures constituted in each depth cue, such as binocular disparity, motion parallax, perspective [Ernst & Banks (2000), Hillis et al. (2004), for a review see Ernst and Bulthoff (2004)].

In some situations, however, the dynamic adjustments cannot reach a resolved state. For example, if the subject is presbiopic, the dynamic accommodative response is unable to achieve focus. Similarly, if the sensory stimulus is manipulated such that the different measures of depth (e.g. disparity and accommodation) are no longer in register, then the visual system is presented with what we have been referring to as cue conflict. In such cases, the visual system may arrive at a perceptual product, but one that is often not stable, or one where the metrics do not conform to information about the object as measured by each cue individually. Conflict situations can arise internally as a natural consequence of visual dysfunction (as in uncorrected vision) or may be generated externally, such as in virtual reality systems that use stereoscopic displays. Either case leads to a less desirable functional state from the perspective of visual perception.

I will conjecture here that coplanar reflectance change presents a similar *unresolved perceptual state*. Whereas cue conflict is instability in metric resolution brought about by a simulated physical conflict in the visual stimulus, representational conflict can be thought of as inherent to the internal perceptual presentation. Representational conflict arises when the structural interpretation does not conform naturally to the ontology of perceptual presentation and the preferred visual states defined by its intrinsic perceptual encoding. I suggest that representational conflict, such as that generated by coplanar reflectance change, creates an inherently non-desirable functional state of perception. These states generate significant *qualitative* effects that have consequences for the design of artifacts and the visuo-motor components of human-computer interfaces.

What is the evidence that such qualitative effects of surface perception exist? I will briefly review a few examples here that show that the preferences designers make when articulating surface structure of designed objects, as well as certain implicit design rules that are often followed in architecture, reveal that the qualitative effects of the perceptual parsing of surface identity has been implicitly acknowledged in design.

Surface articulation in design

Based on what we have discussed above one might propose three specific aspects of surface articulation that would be critical for diminishing perceptual conflict in surface configurations. They are:

(1) *Avoiding coplanar reflectance change.* In Figure 12a we see an object with a central patch that represents coplanar reflectance change. A suitable design resolution would be to either avoid such differential pigmentation altogether, or provide a structural change at the location where the reflectance discontinuity occurs, if it has to, such as shown in Figure 12b.

(2) *Articulation at surface discontinuities.* As mentioned previously, the identity and structural resolution of surfaces that share a common edge (as in Figure 11) are in conflict. Such conflict might involve surfaces that abut such that their faces are coplanar, or as in Figure 11 where they are not. To ameliorate this conflict one would need to enhance the structural demarcation at the surface discontinuity. This could be done by creating a visible break or displacement, or by enhancing the discontinuity edge. For example, a solution to the surface discontinuity in Figure 11 would be what is shown in Figure 13a and b, where the surfaces become physically displaced, or where they are physically separated. A discontinuity could also be enhanced by a thin “bead” or “ridge” along the discontinuity. A similar articulation is shown in Figure 12c, where the inner edge of the outer surface is provided a ridge.

(3) *Boundary enhancement.* I have stated previously that even in the absence of coplanar reflectance change, or even for a single isolated surface of uniform reflectance, the conflict between the ontological and structural parsing in the perceptual presentation persists. For example, I have suggested that isolated surfaces in the sensorium do not have an infinitely-resolved structural identity within the perceptual presentation that separates them from the “ground” (or background). Thus, even in the case of a single isolated surface, structural identity can be increased by simply enhancing the boundary. That enhancement can be accom-

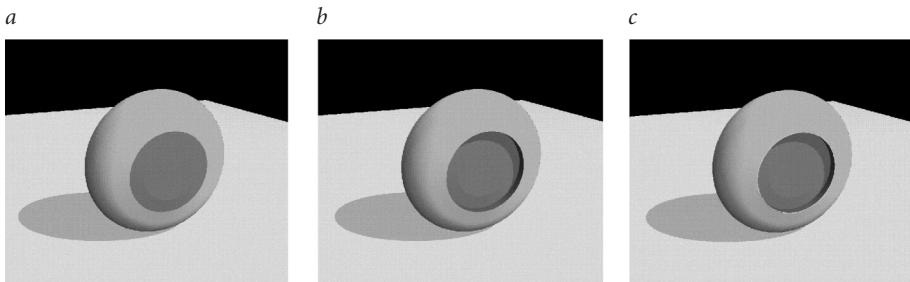


Figure 12. (a) an object with coplanar reflectance change (b) the same object with the surfaces articulated such that the reflectance change is no longer coplanar. (c) Same as (b), except with the addition of a thin demarcating bead along the perimeter of the main surface discontinuity to perceptually enhance the discontinuity.

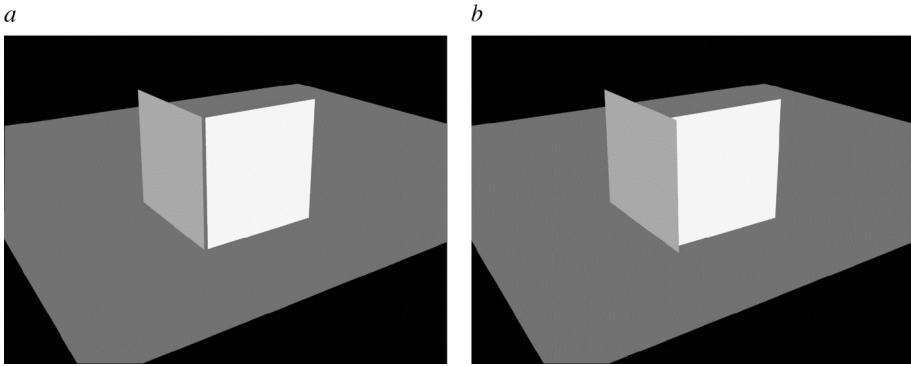


Figure 13. Two ways to enhance the perceptual presentation of the structural identity of surfaces. In (a) the surfaces are separated so that they do not touch. In (b) one of the surfaces is displaced so that it's surface is no longer aligned with the other surface's edge.

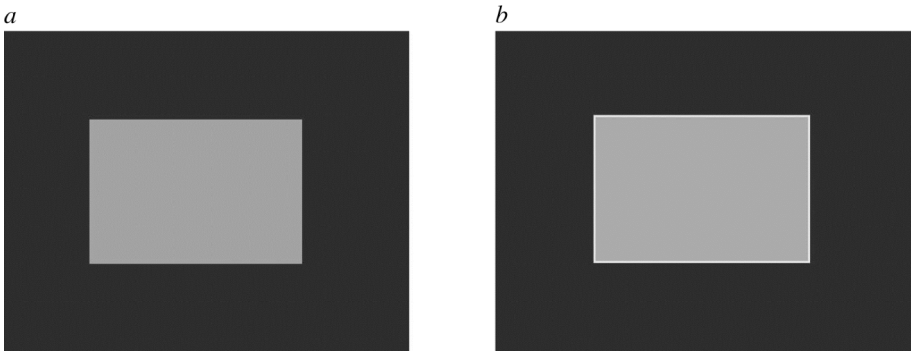


Figure 14. Enhancing the perceptual presentation of the structural identity of an isolated surface. (b) represents a single isolated surface against a black background. I have claimed that even such surfaces pose a certain degree of conflict in the perceptual presentation. (b) enhancing the boundary by providing a thin outline or bead will enhance the structural identity of the surface, and reduce the perceptual conflict.

plished by providing an “outline”; a ridge or bead along the bounding contour. For example, we would predict that the surface in Figure 14b should appear – at a qualitative perceptual level – to be superior to Figure 14a.

Some examples from design

Absent any empirical evidence that the kinds of surface elaborations we have suggested above will indeed generate qualitatively better perceptual presentations, we briefly review some compelling evidence that these kinds of surface elaborations

tions are already codified in written and unwritten design rules. The examples provided here are only a suggestion of the extensive but consistent choices that designers make in configuring surfaces; both in architecture and the design of artifacts. We will go through a few examples of the use of the three methods outlined above. The reader is however encouraged to make careful observations of designed objects to find evidence of these qualitative effects in surface articulation, particularly in examples that follow “classical” rules, as well as those that are considered by consensus to be “good design”. Many of the interesting examples are evidenced in surface-interface details are articulated in architecture, both traditional and modern.

(1) *Avoiding coplanar reflectance change.* In both traditional and modern design, designers have typically eschewed using coplanar patterns as a mode of articulation for visual interest. Articulation in much of classical architecture has usually involved actual change in surface geometry, rather than merely patterned changes in coloration or reflectance. What is particularly avoided are high contrast changes whether or not they are accompanied by an actual change in surface geometry. A example of the use of surface geometry alone for articulation, and a complete avoidance of any change in overall reflectances, is the main façade and interior of Palladio’s San Giorgio Maggiore. In both classical and modern architecture, when reflectance change does occur due to change in material, designers usually go to great lengths to ensure that the reflectance change due to change in material is not coplanar, by carefully disrupting the coplanarity (Figure 12b) or providing an intervening element such as a gap or ridge to break the coplanarity (Figure 12c). Generally, in high design, from architecture, to products or fabrics, creating articulations with patterns of pigmentation has often been considered gauche. For example, one period of architecture that made abundant use of differential pigmentation and patterning without corresponding change in surface structure, particularly in interiors, is Baroque architecture. Though some truly magnificent works have come out of the baroque architectural movement, the term itself has attained a mildly derogatory meaning – to imply overblown or excessive detailing – no doubt because the inconsiderate use of such differential pigmentation techniques in lesser hands produced perceptually poor designs.

(2) *Articulation of surface discontinuities.* One ubiquitous surface discontinuity in architecture is the corner where the walls and ceiling of a space meet. Standard construction treatment of this juncture (in plaster or drywall construction) is to apply a wood molding. This treatment serves a primary purpose of hiding imperfection in workmanship of the finishing of the corner (the hardest part to get right!); but it also inadvertently enhances the demarcation of the discontinuity between two abutting surfaces. However, in more quality construction, it is often

the case that a more elaborate and effective method is commonly applied. This is shown in Figure 15.

Here, the ceiling molding, instead of abutting the ceiling and wall surface, is applied such that a gap between wall and ceiling is implied. Note that this treatment is more tedious than the standard approach, since it requires the carpenter to place the molding quite carefully, making sure the gap has a consistent width throughout the perimeter. However, it has the vivid perceptual effect of making the ceiling surface appear to continue indefinitely past the wall surfaces, and providing it with a much clearer structural identity. In some cases, the molding itself is designed such that the gap is simulated by the nature of the molding's cross section (see Figure 16).

In this case, the effect is not as strong as the gap (Figure 15), but construction is easier since the carpenter can just abut the molding to both surfaces. In modern architecture, where moldings are considered too "traditional" looking, a standard technique is to create a reveal or actual gap between the ceiling and the walls such that they appear as to never meet (see Figure 17a and b).

The approach is more feasible in modern stud and drywall construction. This method of creating a reveal also occurs in prefabricated hung ceilings, wherein the more expensive models of the ceiling grid system will usually have a cross section that creates a physical break between the ceiling surface and wall, and also between the individual panels of the ceiling grid, enhancing the perceptual identity of the ceiling with respect to the wall and the individual panels with respect to each other. More generally, reveals are often provided at any juncture of two materials that



Figure 15. Typical example of ceiling construction articulation that is designed to enhance the perceptual discontinuity between two adjoining surfaces. The wood molding is applied such that its top surface does not abut the ceiling surface, producing a clear perceptual break between the implied wall surface and the ceiling surface; the ceiling appears to continue past the wall.

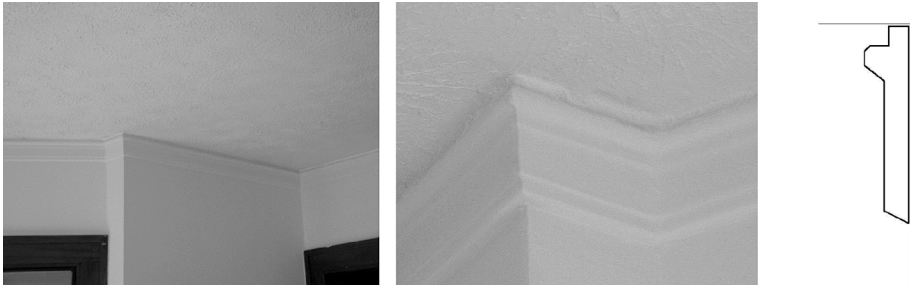


Figure 16. Another example of ceiling construction articulation. Here, the perceptual break is generated by the nature of the ceiling molding itself (cross section shown in right-most panel). This application is a compromise between ease of construction and the clarity of the perceptual break. Since the molding abuts both ceiling and wall surface it is easy to apply, but the perceptual break is not as clear as in Figure 15.

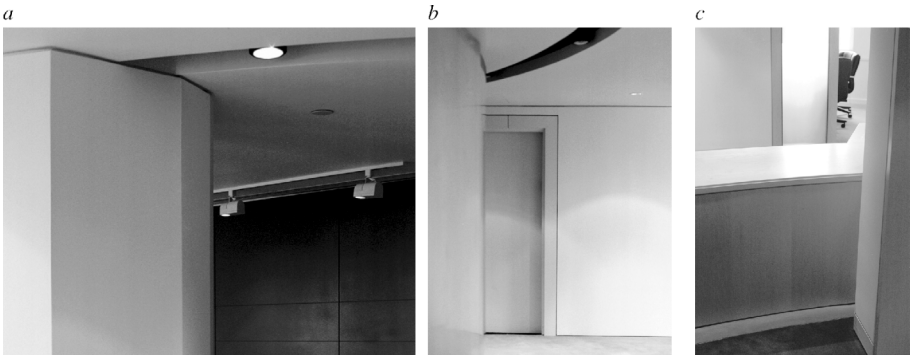


Figure 17. Three examples of the use of gaps or “reveals” used to enhance the perceptual break between surfaces in architectural design. In (a) there is a break between the wall surfaces and ceiling surfaces such that the ceiling appears to continue above the wall. Note that in actuality it does not; the gap is only about an inch deep. (b) shows the same ceiling detail, and a similar reveal detail created at the wall and carpet juncture. Additionally, the coplanar surfaces of the elevator door frame and the wall are also broken with a reveal. (c) shows typical millwork detailing in contemporary interior design, where reveals are created at most junctions of surfaces of similar and different materials, particularly if the materials have significantly different reflectance/coloration.

have different texture or reflectance as shown in Figure 17c where various natural wood, painted wood, and drywall surfaces abut.

A excellent example of articulation of surface discontinuities in product design is the Apple IBook[®] laptop computer. Computer laptops usually consist of two “flaps” one constituting the base and the other the LCD display screen. Standard computer laptops, such as the Dell Dimension[®] shown in Figure 18a, have the

two components attached by a hinge, such that the inner surfaces appear to be connected along the hinged edge. The Apple IBook® (Figure 18b and c), however is designed so that the two faces do not share a common edge when the laptop is opened, enhancing the perceptual identity of the two surfaces and producing a very perceptually satisfying effect. It is interesting to note that the original iteration of this design had a hinge treatment like that of the Dell, and it was only in the later enhanced versions that have the “separated” surfaces.

(3) *Boundary enhancement.* There are some situations where coplanar reflectance change cannot be avoided. One such case is in the design of computer interfaces for software programs or websites. Here, there is a need to partition the display area in order for the user to be able to visually identify different locations of the screen that provide different functionality; for example, by creating rectangular regions, windows, or buttons. Naturally, the only way to demarcate regions on a computer display is by making them appear a different color or lightness from the background. From what we have argued in this paper, each of the regions of different lightness or color is represented in the brain as a distinct surface, yet due to their presentation on a single flat surface, the structural presentation – they are infinitesimally coplanar – generates a perceptual conflict. It is possible that not only

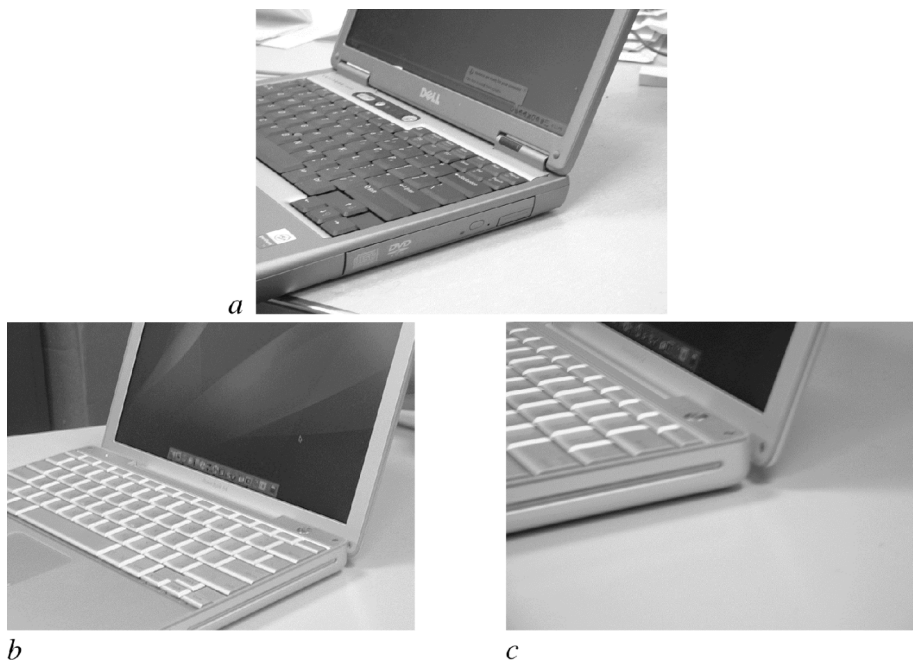


Figure 18.

qualitative visual effects – whether the interface is visually satisfying – but also *allocation of attention* may be disrupted or modulated by this conflict.³³ A standard enhancement to surface identity is usually provided by boundary elements similar in vein to what was suggested in Figure 14b. In graphics this would be a simple outline around the region. With the advent of sophisticated graphics, most interfaces now explicitly use effects that demarcate buttons and windows as being different surfaces overlaying the desktop; effects such as introduction of shadows, edge articulation, etc. It appears that that interface design has naturally evolved to the use of more elaborate methods of surface discontinuity enhancement in order to ameliorate the negative qualitative effects of coplanar reflectance change; and it's deficiencies in sufficiently demarcating regions at the perceptual level due to representational conflict.

There is a more general reason that surfaces may be provided with boundary enhancement, because – as suggested earlier – there may be a conflict in the very identity of the foreground surface with respect to the perceptual ground. Thus a square patch in a dark room has a degree of perceptual conflict that we should then be able to reduce by enhancing its boundary; for example, by providing a fine ridge or lip around its periphery (Figure 14b). Indeed, such methods are used in architecture to make the façade of a building, or the recess of an archway on a building façade stand out.

Conclusions about surfaces and design

Figure 19 shows a view of a house in Rochester, NY. Though it is a non-descript example of turn-of-the-century American architecture, it exhibits a multitude of the surface articulation features we have just mentioned. This house was built in an era when written and unwritten design rules regarding surface articulation were abided to – quite strictly – in quality construction. One can see many examples that we have already spoken about, e.g., the thin wood trim around the window casings that provide the demarcating bead; the detail at the meeting of the brick column surface and roof fascia, which provide an excellent perceptual break between the two surfaces of different reflectance. Another feature is the treatment of the main roof overhang on the front façade. The deep recess of the soffit, and the shadow line it creates, provides a strong perceptual break between the front surface (fascia) of the overhang and the main surface of the buildings façade. The overall façade of the building is further enhanced by the protruding edging at the very top edge of the overhang fascia.

The architectural, design, and computer display examples suggest the following general rule: create a stimulus that results in a non-conflicting presentation of structurally distinct surfaces of similar or different lightness – i.e. one con-



Figure 19.

sistent with the natural ontological presentation of surfaces in perception. This seems central to creating design configurations that are both visually pleasing to the eye, and superior in terms of visual function; which may in turn affect attentional deployment, visuo-motor and cognitive performance. These latter domains may provide opportunities for an empirical and neurophysiological exploration of these qualitative effects in surface perception.

Notes

1. In inferential theories, “cues” are often thought of as objective and prior to the perceptual presentation. The use of the term “cue” in this paper will not imply an objective snippet of information provided for free by some external objective world (except when used to describe inferential theories of perception). Instead, the usage of the term cue in this paper will always

imply a particular mode of measurement of the sensorium that has a necessary dependency on the perceptual presentation. Also see Note 27 and Vishwanath (2005).

2. Avrum Stroll's text is probably the most extensive philosophical analysis on the perception of surfaces. He has analyzed in much detail the epistemological and ontological issues surrounding the notion of surfaces, and many of the issues he raises are certainly relevant to the discussion here. However, I will not attempt to dissect his proposals – supportive or otherwise – since his text takes a particular path in its analytic development; a path which will bring us into philosophical territory that is outside the scope of this chapter. Here, I will try to stay close to the empirical and theoretical results within perceptual science.

3. See Vishwanath (2005).

4. See Leyton (1992) and Albertazzi (this volume). The importance of phenomenological aspects also has a long history in the color perception literature; particularly with Hering (see Turner 1994). The phenomenology of color has also been used extensively in arguing certain philosophical positions on perceptual epistemology and metaphysics (e.g., see Byrne & Hilbert 1997).

5. Some might argue that the 3D percept from a picture is not as convincing as a real 3D object. A great counter example is the stereokinetic effect (Benussi 1911); see Albertazzi (2004) and Zanforlin (1988) for analyses on this phenomena. Also, note that the metrics of a pictorial percept may not match that of real object, and there may be systematic errors in the judgment of the depicted object (Koenderink & van Doorn 2003). What I am pointing out here is that the parsing into *any kind* of 3D structure, including merely the perception of a depth ordering, is entirely inconsistent with the sensory information that signals a planar surface.

6. Leyton (1992) has described how informational and phenomenological aspects such as continuity can be captured within a computational model of perception. In contrast, standard computational approaches (e.g. Bayesian inference) use the notion of continuity or connectivity as an assumption (or prior) along the way to recovering surface orientation. But this application of built-in assumptions has the curious effect of making attributes like continuity an internal property of the inferential device, rather than being part of the information structure of the percept itself (see Vishwanath 2005).

7. Lightness is the intrinsic reflective property of a surface – as we perceive it – that makes it appear light or dark independent of the actual quantity of light emanating from it. Thus we perceive a white surface as being very “light”, even when viewed in very dim conditions where the actual amount of light arriving at the eye per unit surface area might be less than that from a black surface in sunlight, which will appear “dark”. This ability for humans to distinguish the intrinsic reflective property of a material despite variations in actual light intensity is called lightness constancy.

8. I will use “lightness” to generically imply chromatic and monochromatic lightness change unless specified otherwise.

9. Direct realists, sense-data theorists, etc., may balk at the mixing of metaphors here – “are we assuming perception has direct access to a real external world or only to entities in sensation?” The treatment here will assume the existence of a structured external world, but one without an objective *description* independent of perception. When I refer to an *r-surface*, I am referring to an entity that in our folk psychology we attribute to the external world, and whose geometry and reflectance properties we can putatively measure within the framework of our visuo-motor sensorium. When I refer to a *p-surface*, I am referring to the perceptual presentation of a surface. I

am claiming here that, essentially, an r-surface is neither a perceptual entity, nor an objective description of an external entity. Instead, it is a component of a descriptive ontology of the external world derived post-perceptually by application of measurement and cognitive inferences.

10. Note that the non-coplanarity suggested in Figure 1c is an exaggeration for illustration purposes; we will later suggest that the actual percept is of five surfaces that are only infinitesimally non-coplanar.

11. r-surface *luminance* is merely the amount of light emanating from an r-surface in a particular direction, and is a function of the r-surface reflectance and the nature of the r-surface illumination. Thus, a dark r-surface under bright illumination may have higher luminance than a light r-surface under dim illumination. The pattern of light created at the retina (the sensory input or *irradiance*) is proportional to r-surface luminance. *Brightness* is the perceived luminance of a p-surface. Note again that p-surface *lightness* (how intrinsically dark or light a p-surface appears) will not necessarily be proportional to the *brightness* alone, since brightness is a function of both perceived lightness and perceived illumination.

12. See Note 11.

13. p-surface lightness is obviously correlated with some physical flux that for certain functional purposes may be called reflectance. Reflectance can be thought of as a property projected onto the world cognitively from available perceptual entities and properties, and our sensory measurements. In other words, the ontological status of lightness is *prior* to any definition of reflectance.

14. I will use the term “structural presentation” to loosely imply that part of the perceptual content that we might think of as spatial geometry. The term *geometry* may not adequately capture the true nature of the structural presentation (see Leyton 2001).

15. Slant is defined as the angle between the surface normal (at a point on the surface) and the line of sight (to that point). A frontoparallel surface has 0° slant. Tilt at a point on a surface is defined as the direction (in the image plane) along which slant is increased by the greatest amount. For a planar surface rotated about a vertical axis, tilt is horizontal or 0.

16. Recall that I am using the term brightness interchangeable for the intensity of light at the retinal image, as well as the perceived surface luminance, i.e., p-surface brightness.

17. We can think of a veridical percept as one where the perceived p-surface shape and lightness is consistent with the various cues that the visual system uses to infer surface properties, and where the perceived p-surface shape, lightness, and brightness is consistent with the measured r-surface geometry, reflectance, and surface luminance.

18. A distinction must be made here between computational models of cue combination that are directed at understanding how we might make judgments or estimates of metric properties, from those directed at understanding how the actual percept of surface shape or reflectance is achieved. The latter entail a categorical presentation that is not properly thought of as a metric judgment, though the literature often conflates the two (Vishwanath 2005). For an excellent review of the use of Bayesian approaches to cue combination for metric judgments, see Ernst and Bulthoff (2004).

19. Note that a simple image-statistics based Bayesian prior for non-coplanar surfaces may be constructed by analyzing a large number of ecological images with range data; a potential distinct surface patch might be simply identified as any continuous region with homogenous luminance. Relative occurrence of coplanar reflectance change can then be identified by using the range data. How this will favor a prior for coplanar reflectance surfaces (P_1) or a prior for dis-

tinct homogeneous reflectance surfaces located at different depths (P_2) will depend on a variety of complex factors such as the scale of analysis, definition of the ecology that is assumed, etc. But even if the resulting prior for distinct homogeneous surfaces (P_2) has a high value, we would still need to explain within a Bayesian maximum likelihood model why the extremely high likelihood ratios for coplanar reflectance change, provided by disparity and motion parallax, get overridden by a prior that most likely has a broad distribution. In order to support a percept of 3D change given coplanar reflectance change under binocular viewing, the prior for coplanar reflectance change on surfaces (P_1) would need to be *zero* in order to counteract the high likelihood ratios signaling a flat surface. Indeed, the conjecture presented here entails that under a Bayesian model, the prior for coplanar reflectance change would have to be zero! However, the claim is not that it is an ecological outcome, but rather an intrinsic presentational constraint of perception.

20. The actual perceptual presentation is ambiguous. An equally valid interpretation would be of a larger dark surface behind, with two lighter flanking surfaces lying in front.

21. Ontologically speaking, we cannot even have the painter hang around the workshop to paint the sheet metal in different, albeit uniform, colors. The color of the sheet metal, is not a secondary property, it is just the outcome of the metal worker choosing an identifiable sheet of metal to use.

22. Observe that when viewing the image in Figure 4, the measures of depth (disparity cues, convergence, accommodative cues, motion parallax, and surface micro texture) are *all* in agreement, and they are all consistent with a planar surface, which *is* the physical stimulus. However they may be said to be in metric conflict with the 3D interpretation. In an inference model such a 3D interpretation will be ascribed to the so-called monocular cues (shading, perspective, symmetry, etc.)

23. Patterned surfaces can be considered to be collections of multiple surfaces measured – via visual and tactile cues – to be coplanar, but in conflict with the perceptual presentation of surface identity. In fact, the terms “surface pattern” or “surface markings” might be considered to be the cognitive label that we apply, by convention, to such collection of infinitesimally coplanar p-surfaces.

24. This natural assumption we make in our naïve ontology of the world, appears to derive from a necessary constraint in perceptual representation.

25. See Note 14.

26. Note, however, that they *do* represent “cue conflict”, because if we compare measures that we might derive from brightness gradients, perspective geometry, and texture gradients with ones that we might derive from disparity, motion parallax and accommodation, we will see that the two sets of measures are in conflict. Indeed the perceived 3D interpretation is in complete conflict with motion parallax and disparity cues.

27. Naturally, pictorial space always constitutes a *cue-conflict* interpretation (see Note 26 above).

28. Stroll (1988) has analyzed at length whether an object with discontinuous surfaces (e.g. a table) is correctly thought of as having a single surface or multiple surfaces. While his conclusions primarily hinge on linguistic usage and functional considerations (e.g. “wipe the surface of the table”), the proposals here are strictly within the domain of perceptual presentation and the nature of its encoding.

29. This could also be thought of as a conflict between two *levels* of presentation categorical and structural (see Albertazzi, this volume).

30. Another way of thinking about this is that the *resolution* of the parsing at the ontological level is infinite, but the resolution of the metric structural representation is never so.
31. Ewald Hering was a strong proponent of the idea of perceptual states being equivalent to equilibrium states of the organism in response to stimulation (see Turner 1994).
32. Strictly speaking, within a proper epistemological model of perception, such optimal combination of sensory information should be thought of as dynamic adjustments or calibrations that achieve internal coherence of the percept. “Cues” are merely a specification of measurements that we can make on the percept, internally via psychophysics, or of the external entity, via device-based measurement. The optimal state is when the psychophysical measures and external measures agree in a statistical sense. This is leaving aside, naturally, the stronger, often apparently non-optimal constraints on interpretation forced by the ontological and representational schema of perception.
33. See Nakayama et al. (1995) for a discussion on attentional deployment and surface perception.

References

- Adelson, E. H., & Pentland, A. P. (1996). The perception of shading and reflectance. In D. Knill & W. Richards (Eds.), *Perception as Bayesian Inference* (pp. 409–423). New York: Cambridge University Press.
- Albertazzi L. (2004). Stereokinetic shapes and their shadows. *Perception*, 33, 1437–1452.
- Allison, R. S., & Howard, I. P. (2000b). Stereopsis with persisting and dynamic textures. *Vision Research*, 40, 3823–3827.
- Anderson, B. L. (1997). A theory of illusory lightness and transparency in monocular and binocular images. *Perception*, 26(4), 419–453.
- Benussi, V. (1911) Über die motive der Scheinkörperlichkeit bei umkehrbaren Zeichnungen. *Archiv für die gesamte Psychologie*, 20, 363–3961.
- Barrow, H. G., & Tannenbaum, J. R. (1981). Interpreting line drawings as three-dimensional surfaces. *Artificial Intelligence*, 17, 75–116.
- Binford, T. O. (1981). Inferring surfaces from images. *Artificial Intelligence*, 17, 205–244.
- Byrne, A., & Hilbert, D. R. (1997). Colors and reflectances. In A. Byrne & D. R. Hilbert (Eds.), *Readings on Color, Volume 1: The Philosophy of Color* (pp. 263–288). Cambridge: MIT Press.
- Ernst, M. O., & Banks, M. S., (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429–433.
- Ernst, M. O., & Bühlhoff, H. H., (2004). Merging the Sense into a Robust Percept. *Trends in Cognitive Sciences* 8(4), 162–169.
- Gibson, J. J. (1950). *The Perception of The Visual World*. Boston: Houghton Mifflin.
- Hills, J. M., Watt, S. J., Landy, M. S., & Banks, M. S., (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, 4, 967–929.
- Julesz, B. (1971). *Foundations of Cyclopean Perception*. Chicago: University of Chicago Press.
- Koenderink, J. J., & van Doorn, A. J. (2003). Pictorial space. In H. Hecht, R. Schwartz, & M. Atherton (Eds.), *Looking into Pictures: An Interdisciplinary Approach to Pictorial Space* (pp. 239–299). Cambridge: MIT Press.
- Leyton, M. (1992). *Symmetry, Causality, Mind*. Cambridge: MIT Press.
- Leyton, M. (2001). *A Generative Theory of Shape*. Heidelberg: Springer-Verlag.

- Marr, D. (1982). *Vision*. San Francisco: Freeman Press.
- Nakayama, K., He, J. Z., & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In D. N. Osherson & S. M. Kosslyn (Eds.), *An Invitation to Cognitive Science, Volume 2: Visual Cognition*. Cambridge: MIT Press.
- Stroll, A. (1988). *Surfaces*. Minneapolis: University of Minnesota Press.
- Turner, R. S. (1994). *In the Eye's Mind: Vision and the Helmholtz-Hering Controversy*. Princeton: Princeton University Press.
- Vishwanath, D. (2005). The epistemological status of vision and its implications for design. *Axiomathes*, 15(3), 399–486.
- Zanforlin, M. (1988). The height of a stereokinetic cone: A quantitative determination of a 3-D effect from 2-D moving patterns without a rigidity assumption. *Psychological Research*, 50, (3), 162–172.

Evidence suggestive of separate visual dynamics in perception and in memory

Timothy L. Hubbard and Jon R. Courtney

PART I: Dynamics and displacements

The appearance of an object often reflects the forces that created or operated on that object, and so even though a work of art such as a painting or sculpture is actually a static object, the shape, location, or orientation of that object could convey dynamic information regarding its creation or the forces that have operated or are operating on it. As discussed below, previous research has found that observers are sensitive to such dynamics, and that memory is often biased or displaced (e.g., the remembered final position of a moving stimulus is mislocalized away from the actual final position) in ways consistent with the influence or application of such dynamics.

Dynamics in everyday life

There are several potential types of dynamic information that could occur in a static work of art such as a painting or a sculpture, and these include dynamics arising from the structure of the object and dynamics arising from physical forces. Examples of the former include pointing and directed tension, and examples of the latter include momentum and gravitational attraction.

Dynamics arising from structure

The shape and structural skeleton of an object influence the dynamics arising from structure. Also, the configuration within which the object is embedded, including the relative location, orientation, and movement of that object relative to other nearby stimuli, influences the dynamics arising from structure.

Pointing. The simplest of the implied dynamics arising from structure involves “pointing,” and examples of this are shown in Figure 1. An equilateral triangle appears to point in only a single direction at a given moment in time (Attneave 1968), but the direction of perceived pointing is biased (a) by the orientation and shape of the surrounding elements (Palmer 1980), (b) toward a direction aligned with or perpendicular to a configural line (Palmer & Bucher 1981), and (c) in a direction parallel with an internal striping, especially when the stripes are relatively broad (Palmer & Bucher 1982). The direction of perceived pointing of a triangle along a given axis is enhanced when motion of the triangle occurs along that axis (Bucher & Palmer 1985), and similarly, the direction in which a target is perceived to point influences the direction of subsequent apparent motion of that target (McBeath & Morikawa 1997; McBeath, Morikawa, & Kaiser 1992). Also, memory for a pointing shape can be displaced in the direction of pointing (Freyd & Pantzer 1995), and if a target moves in the direction of its point, forward displacement is larger than if the target moves in another direction (Freyd & Pantzer 1995; Nagai & Yagi 2001).

Shape and directed tension. Arnheim (1974) suggested that visual perception consisted in the experiencing of visual forces. However, a painting or sculpture is an immobile object, and so how might such an object possess “visual force”? Many natural objects possess visual forces (i.e., dynamics) because their shapes retain traces of the physical forces that created those objects, and examples of this can be found in the winding or twisting shapes of tree trunks, the traces of waves on a beach, and the spiral curve of a snail’s shell. Indeed, the groundbreaking work of Thompson (1942/1992) highlights that the shape and form of many biological structures can be viewed as a “diagram of the forces” that have operated on those structures. Such an encoding of dynamic information is not limited to natural objects, but can also be found in many artifacts. Along these lines, Arnheim suggests handwriting preserves a record of the motions of the hand that created that writing, and this is consistent with subsequent findings that observers presented with Chinese character-like figures are able to use dynamic information within those stimuli (e.g., thin connecting lines between major strokes) to determine the temporal order in which the lines comprising that figure had been drawn (e.g., Babcock & Freyd 1988; Tse & Cavanagh 2000).¹

The dynamics of a given shape are influenced by the pattern of directed tension specified by that shape, and both Arnheim (1974) and Leyton (1992) suggest that perception of this tension is an integral part of the perceptual experience of the shape. One way in which directed tension can be produced is by deformation of an object away from a shape that is relatively regular, simple, and symmetrical; indeed, the more a given object is deformed away from a regular, simple, and symmetrical shape, the greater the directed tension toward a more regular, simple, and symmetrical shape. The directed tension results in a striving of an object

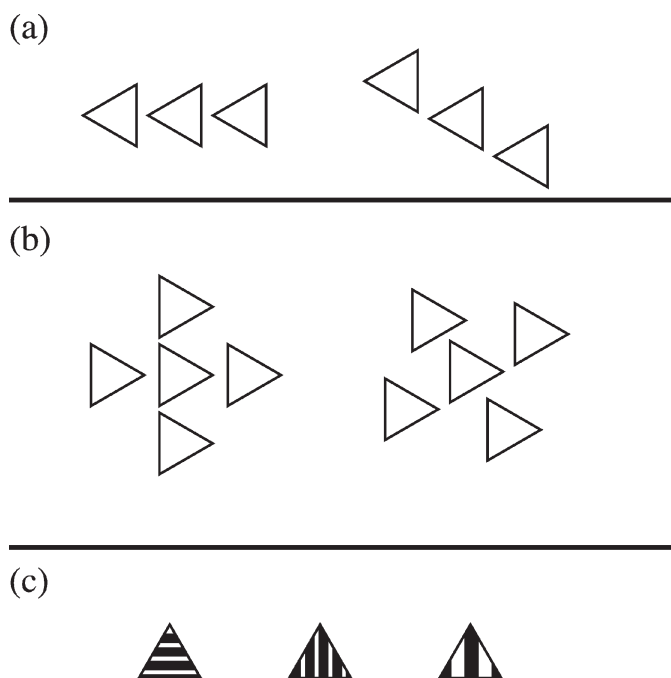


Figure 1. Examples of the dynamics of pointing. In panel (a), the effects of orientation and shape of surrounding elements on the direction of perceived pointing are shown; the stimuli on the left are more likely to be perceived as pointing toward the left, and the stimuli on the right are more likely to be perceived as pointing toward the upper right. In panel (b), the effects of alignment with a configural line are shown; the stimuli on the left are more likely to be perceived as pointing toward the right, and the stimuli on the right are more likely to be perceived as pointing toward the upper left. In panel (c), the degree to which the triangle points upward is lowest on the left and highest on the right. Adapted from Palmer (1980) and Palmer and Bucher (1981, 1982).

back toward a more regular, simple, and symmetrical shape, and such a striving is consistent with the Gestalt law of *Pragnanz*. Because visual forces and directed tensions reflect a change from an initially more symmetrical object, Leyton suggested that any deviation of an object from symmetry can be seen as the operation of forces upon that object. Thus, when we view the shape of an object, we are actually viewing a memory or record of forces that acted upon that object, and the task of perception is to recover that history (for a more formal and rigorous treatment of these ideas, see Leyton 1992).

The dynamics of shape and directed tension also reflect the structural skeleton of a shape. The structural skeleton consists primarily of the axes and principle lines of a shape, and examples of the forces related to the structural skeleton are shown in Figure 2. In the case of a circle, there is a dynamic along the radii. Arn-

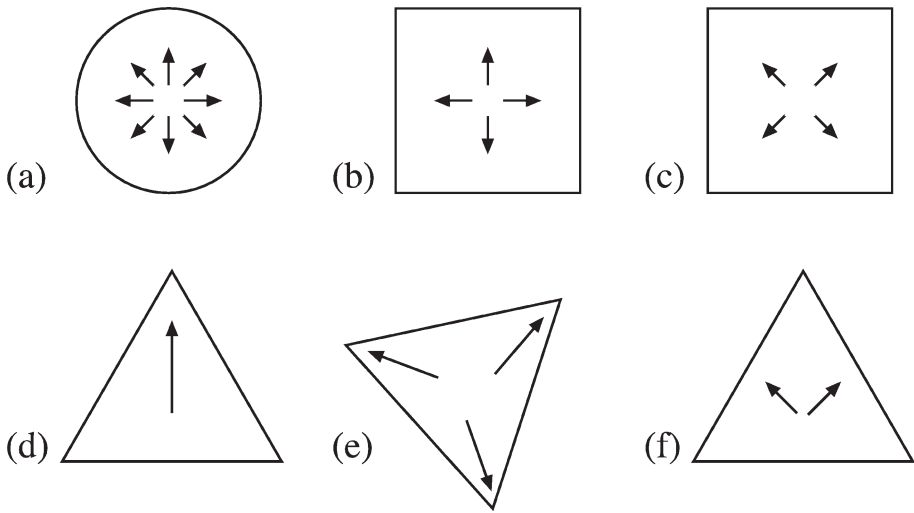


Figure 2. Examples of directed tension arising from shape. In panel (a), a circle exhibits directed tension along the radii. In panels (b) and (c), a square exhibits directed tension along the horizontal and vertical axes of symmetry and along the diagonals. In panel (d), an upward moving triangle exhibits directed tension in the direction of motion. In panel (e), a stationary triangle in which none of the sides is aligned with the horizontal or vertical exhibits directed tension toward each of the points. In panel (f), a stationary triangle in which one side is aligned with the horizontal exhibits directed tension outward against the sloping sides. Adapted from Arnheim (1974).

heim argues this dynamic is clearly shown by the illusion of gamma motion, in which a circular target that suddenly (a) appears is perceived to expand from the center outward, or (b) disappears is perceived to contract from the periphery inward (e.g., see Bartley & Wilkinson 1953; Harrower 1929; Winters 1964). Squares and rectangles exhibit similar dynamics outward along their axes of symmetry and along their diagonals. Consistent with earlier observations, a triangle has a dynamic in the direction of its point(s), and this dynamic is presumably related to observations that a triangle “points”. Arnheim also suggested that if one side of an equilateral triangle is horizontal and that triangle points upward, then the two remaining sides thrust upward and outward as if hinged at the apex (but see Hubbard & Blessum 2001); however, if none of the sides of an equilateral triangle are horizontal, then directed tension is toward all corners.

The place of a given shape within the larger configuration of the scene, and any movement of a given shape relative to other elements in the scene, can contribute to the directed tensions of a given shape. Arnheim (1974) points out how a windmill in which the arms are perfectly horizontal and vertical does not exhibit a dynamic of rotation, and this occurs because such a configuration is a balanced

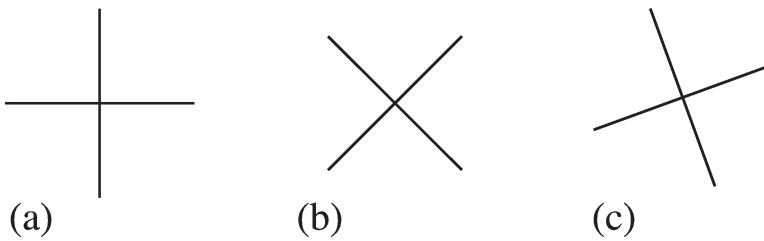


Figure 3. A schematic of the dynamics of the arms of a windmill. In panel (a), the arms are aligned horizontally and vertically, and the symmetry and balance result in little sense of rotation. In panel (b), the arms are aligned with the diagonals, and the symmetry and balance result in relatively little sense of rotation, although the sense of rotation is sometimes slightly larger than in panel (a). In panel (c), the arms are unbalanced and asymmetrical about the horizontal and vertical, and there is a much greater sense of rotation than in panels (a) and (b). Adapted from Arnheim (1974).

and static position. However, as shown in Figure 3, if the arms are at some intermediate orientation (i.e., if the arms are not aligned with the horizontal, vertical, or diagonal), then a dynamic of rotation is much more strongly perceived. In the case of a windmill, the relationships between the individual arms in isolation are exactly the same when the arms are horizontal and vertical as when the arms are at intermediate orientations, but in the former configuration a dynamic of rotation is relatively weak, and in the latter configuration a dynamic of rotation is relatively strong. The arising of a dynamic of rotation out of the configuration of the arms is consistent with the example given earlier of how the configuration of a set of triangles could determine the direction in which those triangles were perceived to point, and is also consistent with a more general Gestalt notion that the context in which a stimulus is embedded influences perception of that stimulus.

Another example of how the location of a shape within a scene can contribute to the dynamics of that scene comes from a consideration of “balance” in pictorial composition. Arnheim (1974) suggested dynamic tension in a work of art can arise from configurations that appear “unbalanced.” For example, paintings are perceived as more unbalanced when they depict more massive objects in the top half of the picture plane (Winner, Dion, Rosenblatt, & Gardner 1987), and it might be possible to account for such directed tension and lack of balance by a consideration of the effect of implied gravity in the scene: The asymmetrical direction of gravitational attraction makes it more likely that a target would move downward rather than upward, and the possible range of potential downward motion would be larger for targets higher in the picture plane than for targets lower in the picture plane. Thus, observers expect an object higher in the picture plane to fall (or fall farther), whereas observers might not expect an object lower in the picture plane to fall (or fall as far). The expectation that a target higher in the picture plane

is more likely to fall (or fall farther) suggests less equilibrium or stability, and so paintings that contain more massive targets in their top half therefore appear more unbalanced.

Dynamics arising from physical forces

In addition to structural dynamics arising from or within a single shape or from the configuration of objects, dynamics can also arise from the physical forces that currently operate on or are implied to operate on an object. Such dynamics include the influence of physical forces such as momentum and gravitational attraction.

Implied momentum. Although stationary shapes can exhibit directed tension because of their structural skeletons, orientation, or configuration, an even stronger dynamic exists in a moving stimulus. Memory for the final position of a previously viewed moving target is often displaced forward (i.e., remembered as farther in the direction of anticipated target motion) of the actual final position of that target. The initial demonstration of this was provided by Freyd and Finke (1984), who presented observers with a target that consisted of three concentric rectangles that implied motion in a consistent clockwise or counterclockwise direction, and that were referred to as *inducing stimuli* (see Figure 4). A probe rectangle was then presented, and the probe was either rotated slightly backward from the orientation of the final inducing stimulus, at the same orientation as the final inducing stimulus, or rotated slightly forward from the orientation of the final inducing stimulus. Observers were more likely to respond *same* to probes rotated slightly forward than to probes rotated slightly backward, and Freyd and Finke suggested this showed that memory for the target was actually displaced forward. They referred to this as *representational momentum* (for reviews, see Hubbard 1995b, 2005).

Representational momentum has been found in numerous studies, and is the most well-studied of the dynamics related to physical principles. In addition to being found with targets that rotate in the picture plane (Freyd & Johnson 1987), representational momentum has been found with targets that translate horizontally or vertically in the picture plane (Hubbard 1990; Hubbard & Bharucha 1988) or rotate in depth (Munger, Solberg, Horrocks, & Preston 1999). The movement of any context within which the target is embedded influences representational momentum (Hubbard 1993; Whitney & Cavanagh 2002), and this is consistent with the broader effects of configuration on dynamics noted earlier. The forward displacement is influenced by several high-level variables including expectations regarding future target motion (Hubbard 1994a; Verfaillie & d'Ydewalle 1991), the presence of implied friction (Hubbard 1995a, 1998), conceptual knowledge regarding target identity (Reed & Vinson 1996; Vinson & Reed 2002), attributions regarding the source of target motion (Hubbard & Favretto 2003; Hubbard

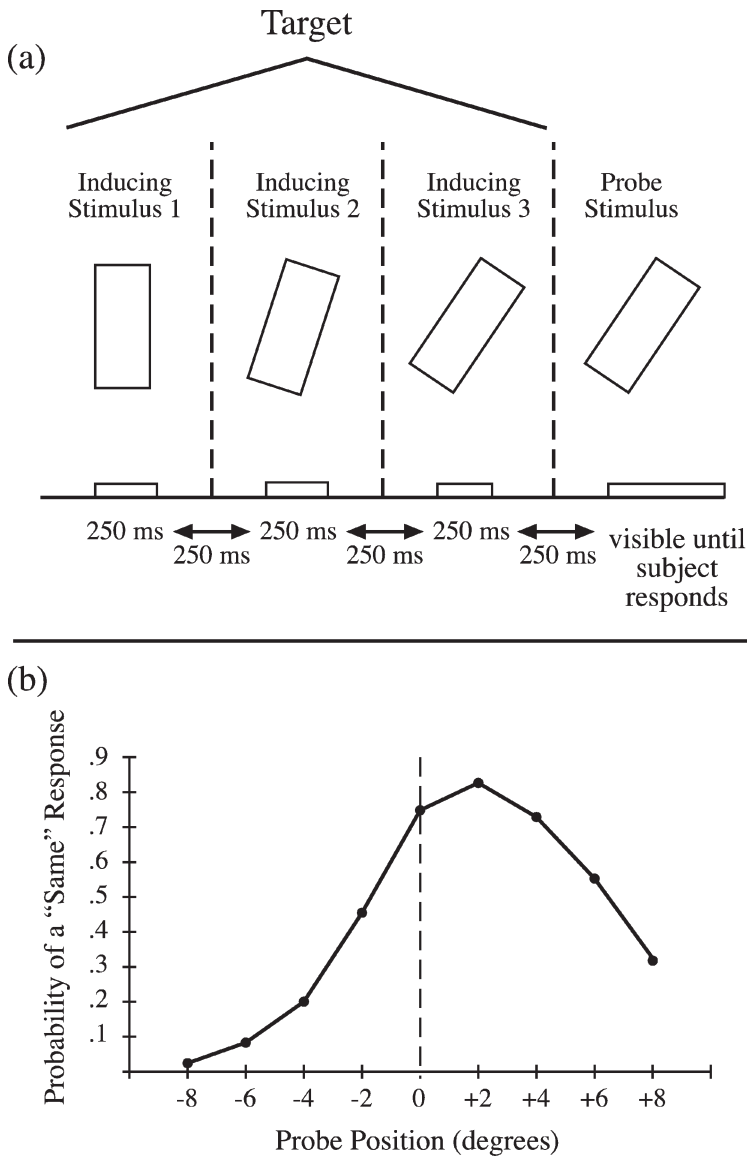


Figure 4. The experimental methodology and results from Freyd and Finke (1984). Panel (a) illustrates a typical trial in which three inducing stimuli (i.e., the target) and a probe are presented. Panel (b) illustrates the probability of a *same* response as a function of probe orientation relative to the final inducing stimulus. The dashed line is the “true same” orientation of the final inducing stimulus, negatively signed probes were rotated backward from the orientation of the final inducing stimulus by the indicated number of degrees, and positively signed probes were rotated forward from the orientation of the final inducing stimulus by the indicated number of degrees. Representational momentum is indicated by the higher probability of a *same* response to positively signed probes.

& Ruppel 2002), and the allocation of attention (Hayes & Freyd 2002; Kerzel 2003), and for the special case of smooth motion, forward displacement is influenced by low-level variables including whether observers visually track the target or fixate a stationary point elsewhere (Kerzel 2000, 2002, 2003).

One of the more interesting findings in the representational momentum literature, and one that is quite relevant to the current consideration, is that it is not necessary to present an actual moving stimulus in order to observe effects of implied motion on displacement in memory. For example, an observer could be shown just a single frame drawn from a previous recording of a longer motion sequence. Examples of such stimuli are shown in Figure 5, and because such stimuli appear to have “frozen” a moment in time, they are referred to as *frozen-action stimuli*. When observers are shown such frozen-action stimuli, and are then shown a probe stimulus from slightly earlier or slightly later in the same motion sequence, they take longer to reject (Freyd 1983) and are more likely to accept as *same* (Futterweit & Beilin 1994) probes from slightly later in that motion sequence than probes from slightly earlier in that motion sequence. Interestingly, paintings or sculptures with high aesthetic value (such as those shown in Figure 6) are often similar to frozen-action stimuli, and conceptualizing of such stimuli as frozen-action stimuli could explain the presence of motion or momentum-like dynamics in objects that are actually stationary, as well as provide a methodology for studying the dynamics of such stimuli.



Figure 5. Examples of frozen-action stimuli. In panel (a), a baseball player is sliding into a base and disrupting the physical stability and equilibrium of the player defending the base. In panel (b), a track and field athlete is executing a pole vault. In panel (c), a small child is jumping off a large toy mushroom. In all three panels, the depicted position is unstable and movement in a specific direction (downward) is implied (i.e., the baseball player defending the base, the pole vaulter, and the child are all suspended in air and must fall back to the ground).

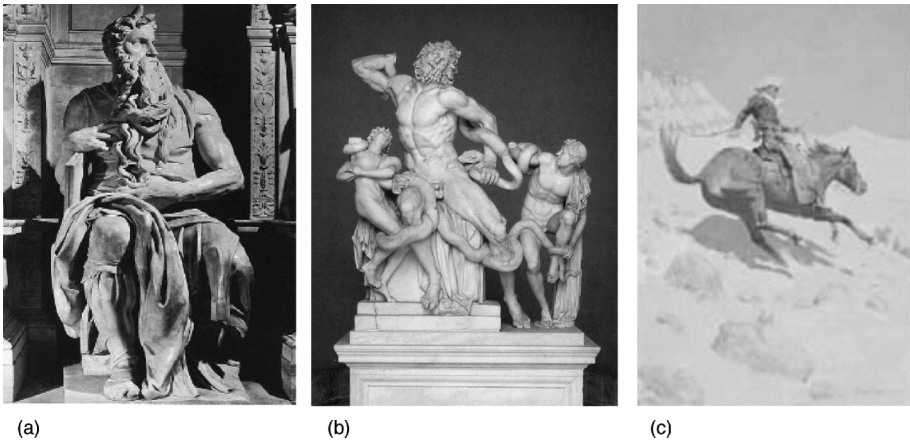


Figure 6. Examples of aesthetic artworks that look like frozen-action stimuli. In panel (a) is a Renaissance sculpture of a seated Moses who appears to be turning toward his left (created by Michelangelo and located in the church of San Pietro in Vincoli in Rome). In panel (b) is the ancient Greek sculpture of Laocoon and his sons battling a sea serpent (created by Athanadoros, Hagesandros, and Polydoros of Rhodes and located in the Vatican Museum in Rome). In panel (c) is a nineteenth century painting of a cowboy riding a horse along a steeply sloping hill (created by Frederic Remington and located in the Amon Carter Museum in Fort Worth, TX). In all three panels, the depicted postures are unstable and movement in a specific direction is implied.

Implied gravity. The example of balance in pictorial composition given above suggests a dynamic related to the direction of gravitational attraction, and such a dynamic has been found in numerous studies. Hubbard and Bharucha (1988; Hubbard 1990) found that memory for the location of a horizontally moving target was displaced slightly downward (as well as displaced forward). Freyd, Pantzer, and Cheng (1988) showed observers a drawing that depicted a flowerpot hanging from a hook or sitting on a table. When observers judged whether a subsequently presented drawing that contained just the flowerpot depicted the flowerpot at the same display coordinates as the original drawing, observers were more likely to respond *same* when the flowerpot in the subsequent drawing was slightly lower in the display. Hubbard (1997) reported the remembered positions of larger targets exhibited larger displacement downward than did the remembered positions of smaller targets, and this occurred regardless of the direction of target motion. Hubbard (2001) reported forward displacement was larger for vertically moving targets lower in the picture plane than for vertically moving targets higher in the picture plane, and this is consistent with the idea that ascending objects decelerate as they rise and descending objects accelerate as they fall. This downward dynamic has been referred to as *representational gravity* (Hubbard 1995b, 1997).

Displacement and aesthetics

Aesthetic response has been suggested to result from dynamic information (e.g., Arnheim 1974), and dynamic information is often explicitly encoded in a work of art. For example, in Japanese *Sei Do* drawing, the force of the depicted object is conveyed in the act of drawing (e.g., a blossoming flower is drawn using gentle motions, whereas a diving eagle is drawn using powerful strokes), and presumably altering the mechanics of the act of drawing to reflect dynamics of the referent object makes such dynamic information more available to an observer. Displacement arising from physical principles such as momentum has similarly been suggested to result from dynamic information (e.g., Freyd 1987, 1992, 1993). Given a common origin in or dependence upon dynamic information, it is possible that aesthetic response and displacement might be related. Aesthetic stimuli could possess more or stronger dynamics, and so memory for those aesthetic stimuli could exhibit more or stronger displacement. It is possible that methods previously used to study displacement and dynamics that are attributable to physical principles might be adapted for the study of aesthetic responses.

Freyd (1992, 1993) suggested implied dynamics such as those arising from the position of a limb in a piece of sculpture or from the relative location of an object depicted in a painting are perceived when an observer first views an element in a work of art, and when an observer then shifts his or her gaze to another element or region of the work of art, memory for the originally fixated element is displaced in the direction of the implied dynamic. If the observer then glances back at the initially fixated element or region, there will be a mismatch between the remembered (displaced) information and the current perceptually sampled information. Such a mismatch could produce a violation of expectancies (an observer would not expect a sculpture or a painted object to have actually moved!) that resulted in a pleasurable (cf. Meyer 1956) or arousing (cf. Berlyne 1971; McMullen 1982) aesthetic experience. The span of time between when fixation moved away from a specific region and then returned to that region might be on the order of a few hundred milliseconds, and if displacement existed only for that brief duration (as suggested by Freyd & Johnson 1987), then it could produce a brief aesthetic experience without permanently biasing memory. Such a suggestion predicts memory for an aesthetic stimulus should be displaced in the direction of the perceived dynamic.

PART II: An empirical study

An aesthetic stimulus whose perceived dynamic properties have previously received considerable attention is the T'ai-chi tu, more commonly known as the

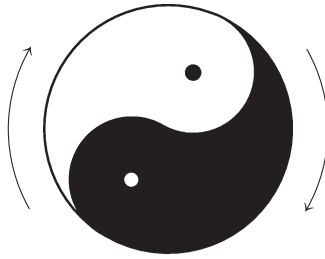


Figure 7. An example of the T'ai-chi tu symbol. The arrows indicate the direction of implied rotation based on Arnheim's (1961) analysis for this specific example.

yin-yang symbol and associated with Taoist philosophy (see Figure 7). Arnheim (1961) presented an extensive analysis of the perceived dynamics of this symbol, and he suggested it exhibits a strong dynamic of rotation. Arnheim (1961) also discussed a number of control stimuli (e.g., a circle divided into two equal parts by a straight line, a set of concentric rings, a spiral) that matched key aspects of the T'ai-chi tu (e.g., equal parts black and white, curved contours), but did not possess the dynamic of rotation possessed by the T'ai-chi tu. Based on Arnheim's (1961) analysis of the T'ai-chi tu and on Freyd's (1992, 1993) speculations regarding dynamics and art, it could be predicted memory for orientation of the T'ai-chi tu should be displaced in the direction of implied rotation, but memory for the orientation of control stimuli that do not possess a strong dynamic of rotation should not exhibit systematic displacement in memory for orientation.

Method

Participants. A total of 84 observers participated in the experiment, and no observer participated in more than one condition. Twenty-four observers viewed the standard T'ai-chi tu (12 viewed clockwise and 12 viewed counterclockwise versions). Twelve observers each viewed the straight-line, concentric rings, and spiral stimuli. Twenty-four observers viewed the monocolour T'ai-chi tu (12 viewed clockwise and 12 viewed counterclockwise versions). Observers were recruited from members of a university community, and were all naïve to the hypotheses.

Stimuli. Examples of the stimuli are shown in Figure 8. The stimuli were 100 pixels (approximately 4.17 degrees of visual angle) in diameter, and were located at the approximate center of the display. On each trial, the stimulus was presented in a target orientation and then in a probe orientation. The probe was at one of five orientations relative to target: -30 , -15 , 0 , $+15$, $+30$ degrees. For the standard T'ai-chi tu and monocolour T'ai-chi tu, negative probe orientations were rotated backward (relative to the direction of implied rotation) from the orientation of the target

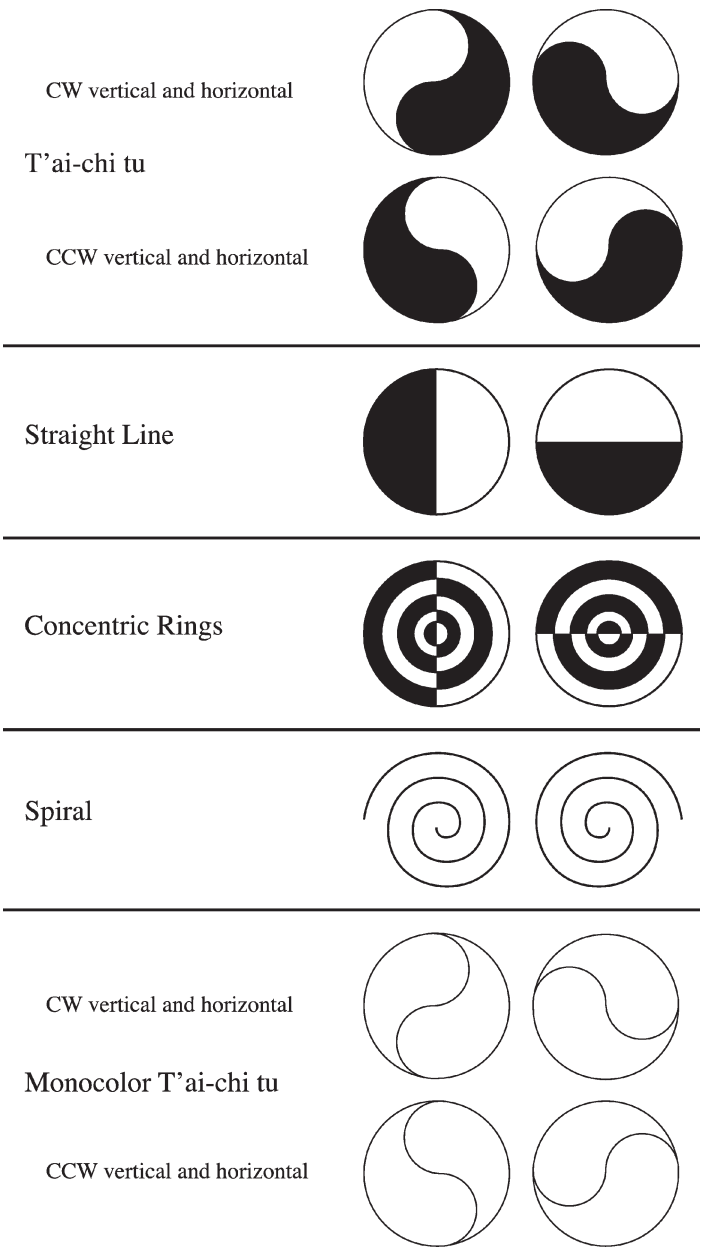


Figure 8. Examples of the standard T'ai-chi tu, straight line, concentric rings, spiral, and monochrome T'ai-chi tu stimuli. Where applicable, vertical and horizontal versions and clockwise and counterclockwise versions of each stimulus are illustrated.

by the indicated number of degrees, and positive probe orientations were rotated forward (relative to the direction of implied rotation) from the orientation of the target by the indicated number of degrees. For the straight line, concentric rings, and spiral stimuli, negative probe orientations were rotated clockwise from the orientation of the target by the indicated number of degrees, and positive probe orientations were rotated counterclockwise from the orientation of the target by the indicated number of degrees. For all stimuli, the zero probe orientation was the same as the orientation of the target. The background was a medium gray. In order to obtain meaningful proportions of *same* responses, each probe was presented 8 times, and each observer received a different random order of trials.

Standard T'ai-chi tu. Stimuli corresponded to a common T'ai-chi tu shape: a circle divided into two parts by an S-shaped line. One part was white and the other part was black; the perimeter of the stimulus was marked by a thin black line. Stimuli were presented in either a horizontal or vertical orientation, and stimuli implied either clockwise or counterclockwise rotation.

Straight line. Stimuli were the same as the standard T'ai-chi tu, with the following exceptions: A straight line along a diameter of the circle, rather than a S-shaped line, separated black and white elements of targets and probes. Stimuli were presented in either a horizontal or vertical orientation.

Concentric rings. Stimuli contained four concentric rings surrounding a circular center; the width of each ring (from the inner edge of that ring to the outer edge of that ring) was 10 pixels (approximately 0.42 degrees of visual angle), and the radius of the circular center was 10 pixels; the total diameter of the stimulus was 100 pixels. Each stimulus was bisected by a horizontal or vertical diameter line which divided the concentric rings into half-rings and the circular center into two semicircles. One half of a bisected ring was white, and the other half of that bisected ring was black (i.e. if the left or top of a ring was white, then the right or bottom of that ring was black; if the left or top of a ring was black, then the right or bottom of that ring was white) and black and white alternated in adjacent half-rings (i.e., if a given half-ring was white, then a half-ring immediately adjacent to the inner or outer edge of that half-ring was black, and if a given half-ring was black, then a half-ring immediately adjacent to the inner or outer edge of that half-ring was white). The half of the circular center adjacent to a white half-ring was black, and the half of the circular center adjacent to a black half-ring was white. The perimeter of the stimulus was marked by a thin black line. Stimuli were presented in either a horizontal or vertical orientation.

Spiral. Stimuli were lines that exhibited a spiral shape that passed through 3.5 turns (1260 total degrees). Stimuli were presented in either a horizontal or vertical orientation, and stimuli implied either clockwise or counterclockwise rotation.²

Monocolor T'ai-chi tu. Stimuli were the same as the standard T'ai-chi tu, with the following exceptions: the interiors of both parts of each stimulus were white, and the perimeter and the S-shaped dividing line of each stimulus were black. Stimuli were presented in either a horizontal or vertical orientation, and stimuli implied either clockwise or counterclockwise rotation.

Procedure

Observers initiated each trial by pressing a designated key. The target immediately appeared, remained visible for 250 milliseconds, and then vanished. The display remained blank during a retention interval of 250 milliseconds, and then the probe appeared and remained visible until the observer responded. The presentation time and retention interval matched the presentation times and retention intervals in previous studies that reported displacement in memory for frozen-action photographs (e.g., Futterweit & Beilin 1994) or implied motion (e.g., Freyd & Finke 1984) stimuli, and so presumably maximized the chance of observing any potential displacement in the remembered orientation of the stimuli. Observers pressed a key marked *S* or a key marked *D* to indicate if the orientation of the probe was the *same* as or *different* from the orientation of the target.

Results

The probability of a *same* response for each probe for each condition is listed in Table 1. Were observers responding accurately, there would be 0% *same* responses on negatively or positively signed probes and 100% *same* responses on 0 degree unsigned probes. Estimates of the direction and magnitude of displacement were determined by calculating the weighted mean (the sum of the products of the proportion of *same* responses and the distance of the probe from the orientation of the target, in degrees, divided by the sum of the proportions of *same* responses) for each observer for each condition (a weighted mean offers a relatively conservative estimate of displacement; see Faust 1990). The sign of a weighted mean indicated the direction of displacement. For stimuli that implied a single direction of rotation, a minus sign indicated backward displacement (in the direction opposite to implied rotation), and a plus sign indicated forward displacement (in the direction of implied rotation). For stimuli that did not imply a single direction of rotation, a minus sign indicated clockwise displacement, and a plus sign indicated counterclockwise displacement. For all stimuli, the absolute value of

Table 1. Probability of a *same* response to T'ai-chi tu and control stimuli.

	Probe Orientation				
	-30	-15	0	+15	+30
Standard T'ai-chi tu					
Horizontal	.10 (.02)	.57 (.05)	.93 (.02)	.28 (.05)	.02 (.01)
Vertical	.08 (.03)	.52 (.06)	.93 (.02)	.35 (.04)	.02 (.01)
Clockwise	.09 (.03)	.54 (.06)	.96 (.01)	.35 (.05)	.02 (.01)
Counterclockwise	.09 (.02)	.56 (.06)	.91 (.02)	.28 (.04)	.02 (.01)
Straight-Line					
Horizontal	.01 (.01)	.01 (.01)	.96 (.03)	.02 (.02)	.00 (.00)
Vertical	.00 (.00)	.00 (.00)	.97 (.02)	.02 (.02)	.01 (.01)
Concentric Rings					
Horizontal	.06 (.06)	.06 (.05)	.94 (.03)	.15 (.07)	.07 (.06)
Vertical	.07 (.06)	.08 (.05)	.96 (.04)	.12 (.07)	.03 (.02)
Spiral					
Clockwise	.06 (.04)	.27 (.07)	.97 (.02)	.21 (.06)	.02 (.01)
Counterclockwise	.05 (.02)	.21 (.06)	.97 (.02)	.27 (.05)	.02 (.01)
Monocolor T'ai-chi tu					
Horizontal	.05 (.02)	.37 (.06)	.93 (.02)	.15 (.03)	.02 (.01)
Vertical	.03 (.02)	.25 (.05)	.92 (.02)	.16 (.04)	.04 (.01)
Clockwise	.06 (.03)	.36 (.06)	.93 (.02)	.12 (.03)	.03 (.01)
Counterclockwise	.02 (.01)	.26 (.05)	.91 (.02)	.19 (.03)	.02 (.01)

Note. Values in parentheses represent standard errors.

a weighted mean indicated the magnitude of displacement (i.e., larger absolute values indicated larger magnitudes of displacement).

Standard T'ai-chi tu. The average displacement for each orientation was compared to zero, and horizontal, $t(23) = -5.43$, $p < .0001$, and vertical, $t(23) = -2.82$, $p < .01$, targets exhibited significant backward displacement. The average displacement for each direction was compared to zero, and clockwise, $t(23) = -3.38$, $p < .003$, and counterclockwise, $t(23) = -4.46$, $p < .001$, targets exhibited significant backward displacement. This backward displacement is reflected in Table 1 in the greater probability of a *same* response for probes rotated backward than for probes rotated forward (especially when -15 and +15 probes are compared).

Straight line, concentric rings, and spiral. The average displacements did not differ from zero.

Monocolor T'ai-chi tu. The average displacement of each orientation was compared to zero, and horizontal targets exhibited significant backward displacement, $t(22) = -3.72$, $p < .001$, and vertical targets exhibited marginally significant backward displacement, $t(22) = -1.60$, $p < .07$. The average displacement for each

direction was compared to zero, and clockwise targets exhibited significant backward displacement, $t(23) = -4.02$, $p < .001$, and counterclockwise targets exhibited marginally significant backward displacement, $t(21) = -1.40$, $p < .08$. This backward displacement is reflected in Table 1 in the greater probability of a *same* response for probes rotated backward than for probes rotated forward (especially when -15 and $+15$ probes are compared).

Discussion

Memory for the orientation of a T'ai-chi tu was displaced backward, and this displacement is in the direction opposite to that originally predicted. One possible explanation is that Arnheim's analysis (upon which the original prediction was based) is incorrect. In order to examine this possibility, a brief experiment was conducted in which 11 observers were presented with T'ai-chi tu stimuli, and those observers rated on a 1 to 5 scale whether each of those stimuli implied clockwise motion, no motion, or counterclockwise motion (1 = strong counterclockwise motion, 2 = weak counterclockwise motion, 3 = no motion, 4 = weak clockwise motion, 5 = strong clockwise motion). Stimuli suggested by Arnheim's analysis to imply counterclockwise motion received significantly lower ratings than did stimuli suggested by Arnheim's analysis to imply clockwise motion, $F(1,10) = 21.64$, $p < .001$. Additionally, whether stimuli were oriented horizontally or vertically, or corresponded to the orientations of -30 , 0 , or $+30$ probes in Experiment 1, did not influence ratings, all F s < 0.24 , p s $> .67$. An explanation of the backward displacement that is based on the hypotheses that Arnheim's analysis was incorrect and that the perceived dynamic was in the direction opposite to that originally predicted can be rejected. The failure to find significant displacement with the straight line, concentric rings, and spiral stimuli rules out other alternative hypotheses for the backward displacement in memory for a T'ai-chi tu. More specifically, the backward displacement was not due to the T'ai-chi tu being divided into two equal-sized high contrast parts (cf. straight line stimuli), the presence of curved internal boundaries (cf. concentric rings stimuli), or the presence of a more general curvilinear dynamic (cf. spiral stimuli). Instead, the backward displacement might have been due to the shape of the main parts. When those parts preserved the shape typical of a T'ai-chi tu, memory for the orientation of the target was significantly displaced, whereas when those parts did not preserve the shape typical of a T'ai-chi tu, memory for the orientation of the target was not significantly displaced. Furthermore, when those parts were clearly different (as in the high contrast standard T'ai-chi tu) there was stronger displacement than when those parts were not as clearly different (as in the low contrast monochrome T'ai-chi tu), and this is consistent with Arnheim's suggestion that in the latter case the parts

might not be perceived as the same shapes (e.g., perhaps instead perceived as an S-shaped line on top of a circular shape).

Even though the displacement attributable to the shape of the T'ai-chi tu was not in the direction predicted, the importance of shape in specifying which stimuli were displaced and which stimuli were not displaced is consistent with Arnheim's (1974, 1988), Freyd's (1992, 1993), and Leyton's (1992) suggestions that shape preserves information regarding implied dynamics. The importance of shape in the current experiment is also generally consistent with previous findings that memory for a stimulus is influenced by the shape of that stimulus (e.g., presence or orientation of symmetry, Finke, Johnson, & Shyi 1988; Large, McMullen, & Hamm 2003; Tyler 2002). Finding displacement in memory for the orientation of a T'ai-chi tu is also noteworthy, because the T'ai-chi tu stimulus is not a frozen-action photograph of a physical object in motion, but is much more abstract than the real world scenes typically used in studies of displacement that presented frozen-action stimuli. Also, although there have been numerous demonstrations of displacement in remembered orientation for targets undergoing rotation (e.g., Freyd & Finke 1984; Munger & Minchew 2002; Munger, Solberg, Horrocks, & Preston 1999), there has not previously been a demonstration of displacement in remembered orientation for a single stationary target.

As noted earlier, one of the dynamics arising from structure is pointing. The main shapes in the standard T'ai-chi tu and monochrome T'ai-chi tu each tapered to a point, and the directions in which those shapes pointed are consistent with the direction of the displacement in memory for the T'ai-chi tu. Thus, the relative orientations of the pointed ends of the two main parts of the T'ai-chi tu symbol could have resulted in a rotary displacement of the entire stimulus (cf. Freyd & Pantzer 1995). Nagai and Yagi (2001) suggest that pointedness could slightly increase forward displacement; however, effects of pointing on forward displacement are typically relatively weak, and it isn't clear from the experimental data why a potentially weak effect of pointing backward appears to dominate the presumably stronger perceived forward dynamic in determining the overall displacement in memory for the T'ai-chi tu. It is doubtful that pointing is the sole explanation for the backward displacement in memory for orientation, but regardless of the ultimate explanation for the backward displacement of a T'ai-chi tu figure, it must also be explained why the direction of displacement in memory appears at odds with reports that a T'ai-chi tu figure is perceived to exhibit a forward dynamic.

PART III: Backward and forward

The difference between a backward displacement in memory of the T'ai-chi tu and a forward dynamic in perception of the T'ai-chi tu might be explained if dynamics

operative during memory were different from dynamics operative during perception. Also, differences between the backward displacement of an aesthetic stimulus like the T'ai-chi tu and the forward displacement of other types of frozen-action stimuli previously reported in the literature might result from the relative strength of the dynamic, with forward displacement resulting from relatively stronger dynamics (as in frozen-action stimuli) and backward displacement resulting from relatively weaker dynamics (as in the T'ai-chi tu).

Separating perception and memory

One possible explanation for the difference between displacement in memory for the T'ai-chi tu in the experiment reported here and dynamics in perception of the T'ai-chi tu reported in Arnheim's analysis involves the nature of the tasks. In the experiment reported here, observers compared a probe with a previously viewed target, and so the task involved memory. In Arnheim's analysis (and in our subsequent empirical validation of that analysis), an observer would attend to a stimulus that was continually available, and so the task involved perception. Also, studies of dynamics in memory for a stimulus typically involve an explicit comparison between two distinct stimuli or between two examples of a single stimulus, whereas analyses of perceived dynamics of a stimulus typically do not involve any explicit comparison of stimuli. Although there is evidence that memory and perception involve similar processing of stimulus quantities (e.g., perceived magnitude and remembered magnitude are both related to physical intensity by a power function, although the exponent of the power function for a given stimulus dimension is usually less for remembered magnitude than for perceived magnitude, see Algom 1992; Hubbard 1994b), whether or not memory and perception involve similar processing of stimulus dynamics has not been extensively studied.

During the course of viewing a T'ai-chi tu, observers might become sensitive to differences that arise between information from the current perception and information from the prior perception (i.e., differences between current perception and recent memory) of a given stimulus. If memory for some part of the target was displaced forward, then if an observer subsequently refixated that previously viewed part, that re-perceived part would actually look like it was behind its remembered position; in other words, the perceived dynamic would be backward. Such a backward dynamic would be opposite to the forward dynamic typically reported in perceptual experience. However, if memory for the target was displaced backward, then not only would observers be more likely to respond *same* to probes rotated backward, but a subsequently presented probe at the same orientation as the previously viewed target would look like it was actually rotated further ahead of the remembered position of that target; in other words, the perceived dynamic would be forward. Indeed, the only way for a probe at the same orientation as

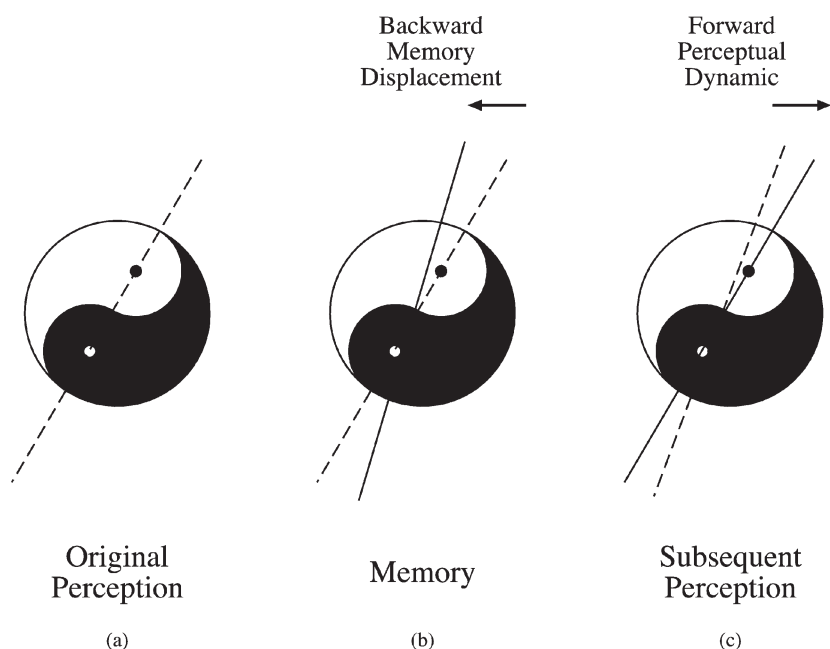


Figure 9. An example using a T'ai-chi tu stimulus of how dynamics in memory and dynamics in perception work in opposite directions. On the left is an initially perceived target, and the perceived orientation is indicated by a solid line. In the middle is the remembered target, and the remembered orientation is indicated by a dashed line. The difference between the dashed line and the solid line indicates the backward displacement in memory. On the right is a subsequent perception of the stimulus, and the difference between the dashed line and the solid line indicates the perceived forward dynamic.

the target (or for a subsequent fixation on the target) to be perceived as having a forward dynamic is for the remembered orientation of the target (to which the subsequent perceived orientation would be compared) to be displaced backward.

How such an explanation is consistent with the backward displacement in memory and the forward dynamic of perception of a T'ai-chi tu is illustrated in Figure 9. In panel (a) is shown the initial fixation of the perceived stimulus, and the perceived orientation is indicated by the solid line. Because this represents the initial perception, there is not yet a remembered orientation. In panel (b) is shown the remembered stimulus at a time after the initial perception, and the remembered orientation from the initial fixation is indicated by a dashed line. The previously perceived orientation from the initial perception is shown as a solid line, and the difference between the orientation of the dashed line and the orientation of the solid line (i.e., the dashed line is rotated backward from the orientation of the solid line) reflects the backward dynamic in memory of orientation. In panel (c) is shown a subsequent perception of the figure, and the perceived orientation

is indicated by a solid line. The remembered orientation from the previous exposure is shown as a dashed line, and the difference between the orientation of the solid line and the orientation of the dashed line (i.e., the solid line is rotated forward from the orientation of the dashed line) reflects the forward dynamic in perception of orientation.

The suggestion that a perceived forward dynamic results from comparison of perceptually sampled information with remembered information that has been displaced backward offers an explanation for a previously puzzling finding. Arnheim (1974) claimed there was a general outward dynamic from the axis of symmetry of an angle, and such a notion suggests memory for the size of an angle should be displaced toward a larger angle size. Hubbard and Blessum (2001) had observers judge whether subsequently presented probe angles were the same angular size as previously viewed target angles, and they found that memory for target angles was actually displaced toward a smaller angle size. This displacement, like the displacement in remembered orientation of a T'ai-chi tu reported here, is in the direction opposite to that initially predicted by Arnheim's aesthetic analyses. The account developed for the T'ai-chi tu can be applied to angle size: Memory for the initially viewed target angle was displaced backward (i.e., inward), but if that target angle remained visible, comparison of current perceptual information with memories from previous fixations would result in the perceived angle seeming to be larger, and so an outward perceptual dynamic would occur.

The directions of displacement

Although a backward dynamic in memory is consistent with the observed displacement in memory for the T'ai-chi tu and a subsequent forward dynamic in perception, why the dynamic in memory is backward is not yet clear. One possibility is that memory exhibits backward displacement because of regression-to-the-mean or a similar averaging toward a central tendency of the stimulus. Such a component has previously been hypothesized in the displacement literature: Freyd and Johnson (1987) reported that forward displacement peaked after several hundred milliseconds and then declined; they suggested the initial forward displacement was due to representational momentum, and that the subsequent decline was due to a bias toward the central tendency of the inducing stimuli (i.e., toward an average of the previous positions of the moving target). In the experiment reported here, there was no actual motion of the target (so representational momentum would not occur), but there still might have been an attempt to recover or extrapolate a possible previous history of the target (cf. Leyton 1992). The existence of any recovered or extrapolated history would allow backward displacement or averaging between the final perceptually sampled stimulus and that history.

Even if backward displacement in memory for a stationary aesthetic stimulus like the T'ai-chi tu can be accounted for, it is not yet clear why displacement in memory for that stimulus is backward, whereas displacement in memory for the object pictured in a frozen-action photograph is usually forward. This is especially troubling given that an aesthetic object (such as those shown in Figure 6) can be viewed as a frozen-action stimulus. One possibility involves the strength of the implied dynamic. In the T'ai-chi tu, the dynamic of rotation is relatively abstract, and possibly weaker than the dynamic of an actual rotating object that is pictured in a frozen-action photograph. It might be that if a forward dynamic is sufficiently strong (as for stimuli pictured in a frozen-action photograph), then displacement is forward in the direction of the implied dynamic, and this would anticipate the subsequent behavior of the pictured object. However, if the forward dynamic is relatively weaker (as for an abstract T'ai-chi tu), then displacement is backward in the direction of a potentially previous state of the stimulus, and this would reconstruct the past of the stimulus. Such reconstruction is consistent with what Leyton (1992) called the primary task of perception: recovering the history of the object.

A forward displacement in the direction of the implied dynamic for only a strong dynamic might still give rise to the potential problem discussed earlier; namely, that an apparent backward dynamic in perception would have to occur because the perceived stimulus would appear to be behind the position suggested by the forward displacement in memory. However, in order for this problem to arise, observers would (a) require extended viewing of the stimulus, and (b) have to assume the stimulus was unchanging. Although both of these conditions can hold for an aesthetic target such as the T'ai-chi tu, neither condition necessarily holds for a stimulus captured in a frozen-action photograph. A frozen-action photograph captures an instant in time out of a longer flow of events, but such a stimulus is a product of developments in photographic technology, and extended viewing of a single instant is not typical of normal perceptual functioning (with the possible exception of eidetic memory). Also, given the relatively stronger dynamics present in frozen-action stimuli, there is no reason for observers to assume such stimuli are unchanging (e.g., observers know a dancer in mid-leap must soon descend). With stronger dynamics, a more adaptive response might be to anticipate the subsequent actions of stimuli, and so displacement is forward in the direction of the implied dynamic.

Fröhlich Effect and Onset Repulsion Effect

With a target such as the T'ai-chi tu or a frozen-action photograph, the target at offset is the same as the target at onset. Perhaps theories regarding displacement in the remembered onset location of a moving target might provide additional explanations or insight regarding why memory for a T'ai-chi tu exhibits backward

displacement, but memory for a frozen-action photograph exhibits forward displacement. Memory for the onset location of a moving target is often displaced in the direction of target motion, and this has been referred to as the *Fröhlich Effect* (e.g., Kerzel & Müsseler 2002; Müsseler & Aschersleben 1998). Forward displacement in the Fröhlich effect for target onset seems similar to forward displacement of representational momentum (Hubbard & Motes 2002). However, memory for the onset location of a moving target can also be displaced backward, and this has been referred to as the *Onset Repulsion Effect* (Thornton 2002). Backward displacement in the Onset Repulsion Effect seems similar to backward displacement of the T'ai-chi tu. An answer to the question of whether memory for the onset location of a moving target exhibits a Fröhlich Effect or an Onset Repulsion Effect might suggest a possible answer to the question of whether memory for an objectively stationary target such as a frozen-action stimulus or an aesthetic object exhibits forward displacement or backward displacement.

One variable that might influence whether a Fröhlich Effect or Onset Repulsion Effect occurs is whether the observer retroactively extrapolates a probable trajectory for the target for the time prior to when the target became visible. When a previous history of the target should have been visible (i.e., the target appeared in the midst of an empty background), an Onset Repulsion Effect occurs, but when a previous history would not have been visible (i.e., the target emerged from behind an occluder), a Fröhlich Effect occurs (see Hubbard & Motes 2005). Observers might be more likely to extrapolate a previous history in the former case than in the latter case, because in the former case observers do not have a reason why the target was not previously visible, whereas in the latter case, observers have a reason why the target was not previously visible (i.e., the target was occluded). The backward displacement of the T'ai-chi tu might be analogous to an Onset Repulsion Effect in which the backward displacement reflects reconstruction of a probable state of the target for the time prior to when the target became visible (cf. Leyton 1992). However, and as noted earlier, if a target has a strong dynamic forward, such as is found in frozen-action photographs, then it might be less important for an observer to extrapolate the past, and more useful for an observer to anticipate the future.

PART IV: Summary and final thoughts

One possible contributor to aesthetic response involves dynamics that previously operated on or are currently operating on the stimulus. A distinction was made between dynamics arising from structure and dynamics arising from implied physical principles, and brief overviews of each type of dynamic were provided. Dynamics arising from structure include influences of pointing, shape and struc-

tural skeleton, and configuration, and dynamics arising from implied physical principles include influences of implied momentum and implied gravitational attraction. A prototypical dynamic stimulus which had previously received extensive aesthetic analysis (the T'ai-chi tu) was presented, and memory for this stimulus was displaced in the direction opposite to the perceived dynamic. This pattern initially seemed inconsistent with previous aesthetic analysis, but could be accounted for when it was realized that memory dynamics and perception dynamics could work in opposite directions. Even so, the reasons why displacement in memory could be backward for some types of stimuli (e.g., T'ai-chi tu) and forward for other types of stimuli (e.g., frozen-action stimuli) are not clear, although a hypothesis introduced here suggests that whether memory is displaced backward or forward depends upon the strength of the dynamic, which might be further indicative of an observer's tendency to extrapolate into the past or into the future.

The ideas discussed here are generally consistent with theories of Arnheim (1974) and Leyton (1992) regarding dynamics resulting from visual forces and directed tensions. The experimental data reported here, coupled with Arnheim's previous aesthetic analysis, suggests memory and perception of a T'ai-chi tu figure were influenced by such dynamics. Indeed, the presence of displacement for a T'ai-chi tu figure and the lack of displacement for the control stimuli are consistent with the hypothesis that stimuli containing more or greater implied dynamics are judged to be more aesthetic. Although suggestive, these results are only preliminary, and a more definitive evaluation of the relationship between dynamics and aesthetics awaits further study (e.g., it is possible that a high level or number of dynamics might be perceived as confusing or chaotic, and so the relationship between aesthetic quality and implied dynamics might be an inverted U-shaped function). The study of the T'ai-chi tu reported here conceived of that aesthetic stimulus as a frozen-action stimulus, and this allowed introduction of a new method for the empirical study of aesthetics. A consideration of dynamics arising from the shape or structure of an object, and how those dynamics might be related to other types of aesthetics, promises to be a fruitful area for further research.

Notes

1. Whether an observer perceives dynamics involved in the previous creation of a stimulus depends in part upon whether that observer was involved in the creation of that stimulus, and this *authorship effect* has been investigated by Knoblich and colleagues. When observers viewed animations of moving dots that recreated dynamics of previously drawn letters or trajectories, those observers could distinguish between animations based on letters or trajectories they previously produced and animations based on letters or trajectories other people previously produced (Knoblich & Prinz 2001) and could distinguish whether a single stroke had been drawn in isolation or followed by another stroke (Knoblich, Seigerschmidt, Flach, & Prinz 2002). Knoblich

and colleagues suggest an observer “reads” the dynamics of a stimulus by activating his or her action representations during perception. There will usually be a better match between one’s action representations and the dynamics of a stimulus if that stimulus had been produced by one’s actions, and so a person is more sensitive to the dynamics of a self-produced stimulus than to the dynamics of an other-produced stimulus.

2. It might be argued that spirals would be perceived as expanding or contracting motion rather than as clockwise or counterclockwise motion. However, expanding or contracting motion along the spiral would move in either a clockwise or counterclockwise direction within the picture plane, and so considering motion of the spiral as clockwise or as counterclockwise can be justified.

References

- Algom, D. (1992). Memory psychophysics: An examination of its perceptual and cognitive prospects. In D. Algom (Ed.), *Psychophysical Approaches to Cognition* (pp. 441–513). New York: North-Holland.
- Arnheim, R. (1961). Perceptual analysis of a cosmological symbol. *Journal of Aesthetics and Art Criticism*, 19, 389–399.
- Arnheim, R. (1974). *Art and Visual Perception: A Psychology of the Creative Eye* (the new version). Berkeley, CA: University of California Press.
- Arnheim, R. (1988). Visual dynamics. *Scientific American*, 76, 585–591.
- Attneave, F. (1968). Triangles as ambiguous figures. *American Journal of Psychology*, 81, 447–453.
- Babcock, M. K., & Freyd, J. J. (1988). Perception of dynamic information in static handwritten forms. *American Journal of Psychology*, 101, 111–130.
- Bartley, S. H., & Wilkinson, F. R. (1953). Some factors in the production of gamma movement. *Journal of Psychology*, 36, 201–206.
- Berlyne, D. E. (1971). *Aesthetics and Psychobiology*. New York: Appleton-Century-Crofts.
- Bucher, N. M., & Palmer, S. E. (1985). Effects of motion on perceived pointing of ambiguous triangles. *Perception & Psychophysics*, 38, 227–236.
- Faust, M. (1990). *Representational Momentum: A Dual Process Perspective*. Unpublished doctoral dissertation, University of Oregon, Eugene, Oregon.
- Finke, R. A., Johnson, M. K., & Shyi, G. C. W. (1988). Memory confusions for real and imagined completions of symmetrical visual patterns. *Memory & Cognition*, 16, 133–137.
- Freyd, J. J. (1983). The mental representation of movement when static stimuli are viewed. *Perception & Psychophysics*, 33, 575–581.
- Freyd, J. J. (1987). Dynamic mental representations. *Psychological Review*, 94, 427–438.
- Freyd, J. J. (1992). Dynamic representations guiding adaptive behavior. In F. Macar, V. Pouthas, & W. J. Friedman (Eds.), *Time, Action, and Cognition: Towards Bridging the Gap* (pp. 309–323). Dordrecht: Kluwer Academic Publishers.
- Freyd, J. J. (1993). Five hunches about perceptual processes and dynamic representations. In D. Meyer & S. Kornblum (Eds.), *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience* (pp. 99–119). Cambridge, MA: MIT Press.
- Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 126–132.

- Freyd, J. J., & Johnson, J. Q. (1987). Probing the time course of representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 259–269.
- Freyd, J. J., & Pantzer, T. M. (1995). Static patterns moving in the mind. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The Creative Cognition Approach* (pp. 181–204). Cambridge, MA: MIT Press.
- Freyd, J. J., Pantzer, T. M., & Cheng, J. L. (1988). Representing statics as forces in equilibrium. *Journal of Experimental Psychology: General*, 117, 395–407.
- Futterweit, L. R., & Beilin, H. (1994). Recognition memory for movement in photographs: A developmental study. *Journal of Experimental Child Psychology*, 57, 163–179.
- Gabrielsson, A. (1999). The performance of music. In D. Deutsch (Ed.), *The Psychology of Music* (2nd ed.) (pp. 501–602). New York: Academic Press.
- Gombrich, E. H. (1960). *Art and Illusion: A Study in the Psychology of Pictorial Representation*. Princeton, NJ: Princeton University Press/Bollingen Series.
- Gombrich, E. H. (1982). *The Image and the Eye: Further Studies of the Psychology of Pictorial Representation*. London: Phaidon Press Ltd.
- Harrower, M. R. (1929). Some experiments of the nature of gamma movement. *Psychologische Forschung*, 13, 55–63.
- Hayes, A. E., & Freyd, J. J. (2002). Representational momentum when attention is divided. *Visual Cognition*, 9, 8–27.
- Hubbard, T. L. (1990). Cognitive representation of linear motion: Possible direction and gravity effects in judged displacement. *Memory & Cognition*, 18, 299–309.
- Hubbard, T. L. (1993). The effects of context on visual representational momentum. *Memory & Cognition*, 21, 103–114.
- Hubbard, T. L. (1994a). Judged displacement: A modular process? *American Journal of Psychology*, 107, 359–373.
- Hubbard, T. L. (1994b). Memory psychophysics. *Psychological Research/Psychologische Forschung*, 56, 237–250.
- Hubbard, T. L. (1995a). Cognitive representation of motion: Evidence for representational friction and gravity analogues. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 241–254.
- Hubbard, T. L. (1995b). Environmental invariants in the representation of motion: Implied dynamics and representational momentum, gravity, friction, and centripetal force. *Psychonomic Bulletin & Review*, 2, 322–338.
- Hubbard, T. L. (1997). Target size and displacement along the axis of implied gravitational attraction: Effects of implied weight and evidence of representational gravity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 1484–1493.
- Hubbard, T. L. (1998). Some effects of representational friction, target size, and memory averaging on memory for vertically moving targets. *Canadian Journal of Experimental Psychology*, 52, 44–49.
- Hubbard, T. L. (2001). The effect of height in the picture plane on the forward displacement of ascending and descending targets. *Canadian Journal of Experimental Psychology*, 55, 325–330.
- Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: A review of the findings. *Psychonomic Bulletin & Review*, 12, 822–851.
- Hubbard, T. L., & Bharucha, J. J. (1988). Judged displacement in apparent vertical and horizontal motion. *Perception & Psychophysics*, 44, 211–221.
- Hubbard, T. L., & Blessum, J. A. (2001). A structural dynamic of form: Displacements in memory for the size of an angle. *Visual Cognition*, 8, 725–749.

- Hubbard, T. L., & Favretto, A. (2003). Naïve impetus and Michotte's "Tool Effect": Evidence from representational momentum. *Psychological Research/Psychologische Forschung*, 67, 134–152.
- Hubbard, T. L., & Motes, M. A. (2002). Does representational momentum reflect a distortion of the length or the endpoint of a trajectory? *Cognition*, 82, B89–B99.
- Hubbard, T. L., & Motes, M. A. (2005). An effect of context on whether memory for initial position exhibits a Fröhlich effect or an onset repulsion effect. *Quarterly Journal of Experimental Psychology*, 58(A), 961–979.
- Hubbard, T. L., & Ruppel, S. E. (2002). A possible role of naïve impetus in Michotte's "Launching Effect": Evidence from representational momentum. *Visual Cognition*, 9, 153–176.
- Kerzel, D. (2000). Eye movements and visible persistence explain the mislocalization of the final position of a moving target. *Vision Research*, 40, 3703–3715.
- Kerzel, D. (2002). The locus of "memory displacement" is at least partially perceptual: Effects of velocity, expectation, friction, memory averaging, and weight. *Perception & Psychophysics*, 64, 680–692.
- Kerzel, D. (2003). Attention maintains mental extrapolation of target position: Irrelevant distractors eliminate forward displacement after implied motion. *Cognition*, 88, 109–131.
- Kerzel, D., & Müsseler, J. (2002). Effect of stimulus material on the Fröhlich illusion. *Vision Research*, 42, 181–189.
- Knoblich, G., & Prinz, W. (2001). Recognition of self-generated actions from kinematic displays of drawing. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 456–465.
- Knoblich, G., Seigerschmidt, E., Flach, R., & Prinz, W. (2002). Authorship effects in the prediction of handwriting strokes: Evidence for action simulation during action perception. *Quarterly Journal of Experimental Psychology*, 55(A), 1027–1046.
- Large, M. E., McMullen, P. A., & Hamm, J. P. (2003). The role of axes of elongation and symmetry in rotated object naming. *Perception & Psychophysics*, 65, 1–19.
- Leyton, M. (1989). Inferring causal history from shape. *Cognitive Science*, 13, 357–387.
- Leyton, M. (1992). *Symmetry, Causality, Mind*. Cambridge, MA: MIT Press.
- McBeath, M. K., & Morikawa, K. (1997). Forward-facing motion biases for rigid and nonrigid biologically likely transformations. *Perceptual and Motor Skills*, 85, 1187–1193.
- McBeath, M. K., Morikawa, K., & Kaiser, M. K. (1992). Perceptual bias for forward-facing motion. *Psychological Science*, 3, 362–367.
- McMullen, P. T. (1982). Connative responses to musical stimuli: A theoretical explanation. *Council for Research in Music Education*, 7, 45–57.
- Meyer, L. B. (1956). *Emotion and Meaning in Music*. Chicago: The University of Chicago Press.
- Munger, M. P., & Minchew, J. H. (2002). Parallels between remembering and predicting an object's location. *Visual Cognition*, 9, 177–194.
- Munger, M. P., Solberg, J. L., Horrocks, K. K., & Preston, A. S. (1999). Representational momentum for rotations in depth: Effects of shading and axis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 157–171.
- Müsseler, J., & Aschersleben, G. (1998). Localizing the first position of a moving stimulus: The Fröhlich effect and an attention-shifting explanation. *Perception & Psychophysics*, 60, 683–695.
- Nagai, M., & Yagi, A. (2001). The pointedness effect on representational momentum. *Memory & Cognition*, 29, 91–99.
- Palmer, S. E. (1980). What makes triangles point: Local and global effects in configurations of ambiguous triangles. *Cognitive Psychology*, 12, 285–305.

- Palmer, S. E., & Bucher, N. M. (1981). Configural effects in perceived pointing of ambiguous triangles. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 88–114.
- Palmer, S. E., & Bucher, N. M. (1982). Textural effects in perceived pointing of ambiguous triangles. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 693–708.
- Reed, C. L., & Vinson, N. G. (1996). Conceptual effects on representational momentum. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 839–850.
- Thompson, D. W. (1992). *On Growth and Form*. New York: Dover Publications (original work published 1942).
- Thornton, I. M. (2002). The onset repulsion effect. *Spatial Vision*, 15, 219–243.
- Tse, P. U., & Cavanagh, P. (2000). Chinese and Americans see opposite apparent motions in a Chinese character. *Cognition*, 74, B27–B32.
- Tyler, C. W. (Ed.). (2002). *Human Symmetry Perception and its Computational Analysis*. Mahwah, NJ: Erlbaum.
- Verfaillie, K., & d'Ydewalle, G. (1991). Representational momentum and event course anticipation in the perception of implied periodical motions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 302–313.
- Vinson, N. G., & Reed, C. L. (2002). Sources of object-specific effects in representational momentum. *Visual Cognition*, 9, 41–65.
- Whitney, D., & Cavanagh, P. (2002). Surrounding motion affects the perceived locations of moving stimuli. *Visual Cognition*, 9, 139–152.
- Winner, E., Dion, J., Rosenblatt, E., & Gardner, H. (1987). Do lateral or vertical reversals affect balance in paintings? *Visual Arts Research*, 13, 1–9.
- Winters, J. J. (1964). Gamma movement: Apparent movement in figural after effects experiments. *Perceptual and Motor Skills*, 19, 819–822.

Illusory space and paradoxical transparency in stereokinetic objects

Mario Zanforlin

Introduction

Pictures are two-dimensional representations of three-dimensional space. The impression of three-dimensionality, namely, perspective three-dimensionality, is combined with a simultaneous impression of a flat surface and is clearly illusory. However, when certain two-dimensional configurations are set in motion they can produce an impression of *real* three-dimensionality quite distinct from the impression of *perspective* three-dimensionality. These are the so called stereokinetic phenomena first observed by V. Benussi and analysed by C. Musatti (1924, 1955). They are produced by various figures drawn or stuck on a larger black disc set in slow rotatory motion (see Appendix).

The stereokinetic objects and their surfaces

When a white disc with a smaller eccentric circle inside it is stuck on a larger black disc, as in Plate 1, Figure 1, and set in motion, a solid cone appears of well defined height in depth and quite undistinguishable from a real object. That is, the surface of the disc between the smaller circle and the larger one is displaced in depth to form the opaque surfaces of the cone. The cone may appear concave or convex and the two appearances alternate during prolonged inspection. When the cone appears convex, it appears solid, but when it appears concave, it appears as empty and the conical surface forms a hole in the larger black disc.

If a second inner circle is added, as in Plate 1, Figure 2, a kind of “lamp shade” appears. One of the circles is displaced in depth away from the observer and captures the surface to form a concave cone. The second inner circle is displaced towards the observer and appears suspended in mid-air. No surface appears to connect this circle to the larger one, so the space inside the cone is part of an open

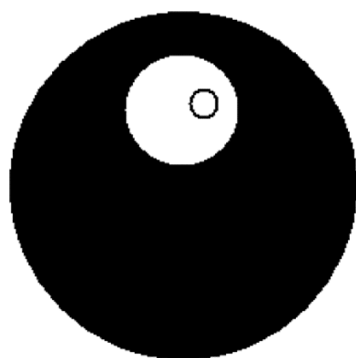


Fig. 1

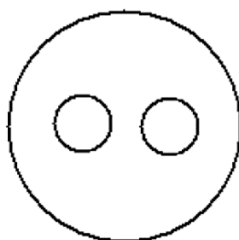


Fig. 2

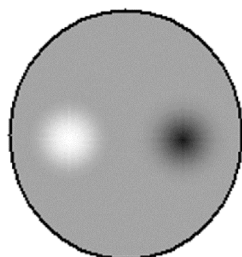


Fig. 3

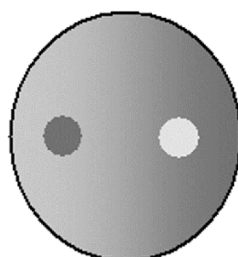


Fig. 4



Fig. 5



Fig. 6

Plate 1. The figures here coloured with shaded or uniform, grey colours, Figures 3–6, can be coloured with any other colour without affecting the apparent transparency described in the text.

space. If this surface could be made visible, a double cone with an enclosed space would be visible.

The fact that this surface is not visible, may be due to the fact that a surface requires only two contours as boundaries, so it can choose one of the two inner circles. Or it may be that, because the circle that forms the concave cone is completely visible, the convex surface formed by the second circle must be transparent and hence not visible.

It is well known that a transparent surface can be perceived as such only if its transparency is not perfect; that is, its colour must differ in some way from that of the object seen through it, or else it must have some shining reflexions.

After various attempts, I found at least two ways of making the convex transparent surface of the double cone visible.

The first was to substitute the two inner circles with two circular gradients of different colours that faded into the uniform colour of the disc surface, as illustrated in Plate 1, Figure 3.

Here a transparent double cone appears. The bottom gradient can be clearly seen through the convex transparent surface. The space enclosed by the two surfaces appears not to be completely empty but filled with rarefied coloured air, like a thin mist.

The second was to colour the entire disc surface with a gradient and add two small discs of uniform colour as in Plate 1, Figure 4. Here a more clearly transparent double cone appears, which looks somewhat like a “whip-top”. This effect can also be obtained with any combination of colour gradients. The enclosed space appears clearer and the colour seems to stick to the surfaces.

One surprising effect of these coloured transparent surfaces is that the colour of the transparent surface does not affect the colour of the disc seen through it. Various combinations of colours have been tested with the same results.

However, as reported earlier (Zanforlin 2003), with different configurations colour gradients are not necessary to make stereokinetic surfaces transparent.

When an oblong figure, as illustrated in Plate 1, Figure 5, is set in rotation, a cylinder appears with a transparent surface through which the small disc at the base of the cylinder appears. Here again the colour of the transparent surface does not affect the colour of the disc seen through it.

The configuration of Plate 1, Figure 6 appears to be of particular interest, where the colour of the cylinder appears more like a kind of light inside the empty space of the cylinder than as a transparent surface; i.e. a coloured light that illuminates the base of the cylinder without affecting its colour.

In all the previous configurations, the inner discs contribute to form and are part of the surface of the transparent objects. However this is not a necessary condition as the inner discs may also appear as suspended inside the transparent object and not part of the surface. This is shown in configuration of Plate 2, Figure 7. Here a solid transparent ellipsoid, an egg, appears inside a vase. The small discs of various colours appear suspended inside the egg and their colours are not affected by the colour of the clearly transparent surface.

Finally, in Plate 2, Figure 8 the blue gradient that fills the vase with the small discs inside seems to be more like a thick mist than a transparent surface.

In all naturally transparent objects, such as glass, water, mist or smoke, the colour of the transparent medium affects the colour of the object seen through it.

In no case does the colour of the transparent stereokinetic surface or misty space, affect the colour of the objects seen through it. So, they appear as purely phenomenal and unnatural, ghostly objects, affected by *paradoxical transparency*.

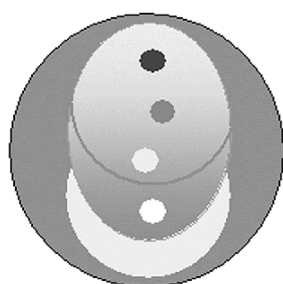


Fig. 7



Fig. 8

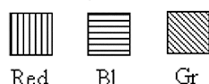
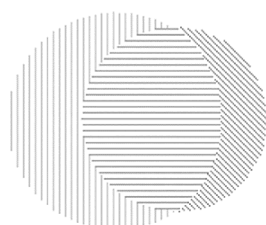


Fig. 9

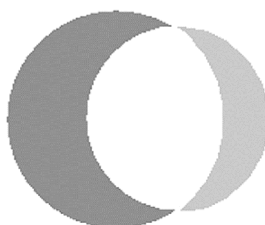


Fig. 10

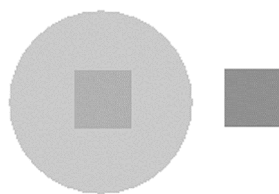


Fig. 11

Plate 2. For Figures 7, 8, 10, and 11, see Plate 1 as regards the colours.

Discussion

The paradoxical transparency here presented is quite different from that first observed by W. Metzger (1956) in stereokinetic objects. Metzger presented two partially super-imposed discs, as illustrated in Plate 2, Figure 9. One disc is red and the other green, while the super-imposed part is blue. It is well known that if we mix red and green we do not get blue. But, when the configuration is set in rotatory motion, one of the discs appears suspended in space and clearly transparent. This is due to the good figural organization of the two discs that enhances the impression of transparency.

However, when the colour of the super-imposed part is quite different from the colour one would expect from the mixture of the colours of the two disc, as in

Plate 2, Figure 10, what appears is not a transparent disc, but a tilted hollow tube of which the superimposed part is the bottom and the two side crescents form the inside and outside walls.

In our case the transparent surface is not partially superimposed but appears to cover the object seen through it completely. The type of transparency, where the transparent surface completely covers the object seen through it, is called Fuchs's transparency (Fuchs 1923) to distinguish it from the case in which the transparent surface covers only partially the object behind it. This type of transparency has been analysed by Masin (1998) who obtained it in two-dimensional configurations, as illustrated in Plate 2, Figure 11. However even in the case of Fuchs's transparency the colour of the superimposed area between the object and the transparent surface is a mixture between the colour of the transparent surface and the colour of the object seen through it. So the colour of stereokinetic transparent surfaces does not affect the colour of the object seen through it. This type of transparency seems to be unique among the various transparencies previously analysed in visual perception. No theoretical or neurological explanation of this phenomenon has so far been advanced.

Acknowledgements

This research has been financed by a grant by the University of Padua. I wish to thank the students and other observers who have contributed to the experimental results here described.

Appendix

The apparatus consists of a metal disc covered by a cardboard black disc 30 cm in diameter, as illustrated in Plate 1, Figure 1. The disc was moved by an electric motor at low speed: 10–20 cycles per minute. The configurations presented were drawn on smaller cardboard discs of 6–8 cm in diameter that could be stuck at any point on the larger black disc.

The laboratory was normally illuminated by daylight and the observers were seated 2 m. in front of the apparatus with their head approximately coaxial to the rotating disc and free to move. (But this distance and the head coaxial are not strictly necessary for perception of the phenomena described, as the latter have also been presented to larger audiences as demonstrations).

The subjects were required to observe the rotating figures with one eye and report what they perceived. All subjects (more than 30) after a few seconds of observation reported perceiving all the phenomena described here.

References

- Fuchs, W. (1923). Experimentelle Untersuchungen über das simultane Hintereinandersehen auf derselben Sehrichtung. *Zeitschrift für Psychologie*, 91, 145–235.
- Masin, S. C. (1998). The luminance conditions of Fuchs's transparency in two-dimensional patterns. *Perception*, 27, 851–859.
- Metzger, W. (1956). Über durchsichtigkeits-erscheinungen. *Rivista di psicologia*, 49, 187–189.
- Musatti, C. L. (1924). Sui fenomeni stereocinetici. *Archivio Italiano di Psicologia*, 3, 105–120.
- Musatti, C. L. (1955). La stereocinesi e il problema della struttura dello spazio visibile. *Rivista di Psicologia*, 49, 3–57.
- Zanforlin, M. (1988). Stereokinetic phenomena as good gestalts. The minimum principle applied to circles and ellipses in rotation: A quantitative analysis and a theoretical discussion. *Gestalt Theory*, 10, 187–214.
- Zanforlin, M. (2000). The various appearances of a rotating ellipse and the minimum principle: a review and an experimental test with non-ambiguous percepts. *Gestalt Theory*, 22, 157–184.
- Zanforlin, M. (2003). Stereokinetic anomalous contours: Demonstrations. *Axiomathes. An International Journal in Ontology and Cognitive Systems*, 13, 389–398.

The neural space of visual shape perception

Charles Connor

One approach to understanding depictive space is to examine directly its physical instantiation in the brain. While the issue of ontological levels is problematic (see Introduction to this volume), the standard scientific view is that neural activity in visual cortex is causally prior to the experiences and reports available through introspection and psychophysical measurement. Neural activity is more complete and in some senses more complex than any verbal reports about it can be. Neural activity may embody an underlying dimensionality inaccessible to introspection but useful in explaining the structure of visual phenomena. For these reasons, direct study of neural activity may illuminate the ground for visual experience in a unique way that complements other approaches.

In humans and non-human primates, visual object information is processed in the ventral visual pathway (Ungerleider & Mishkin 1982), which originates in primary visual cortex (V1) and proceeds through areas V2, V4, and then through multiple stages in ventral occipital/temporal cortex, which in humans include areas V8, LO (lateral occipital), and parts of the fusiform and parahippocampal regions (Epstein et al. 1999; Kanwisher et al. 1997b; Kanwisher et al. 1997a; Kourtzi & Kanwisher 2000; Lerner et al. 2001). Overall levels of neural activity in these human brain regions can be studied using functional magnetic resonance imaging (fMRI). However, the dimensionality of visual information can only be studied at the level of discrete electrical signals (spikes) emitted by individual neurons – these are the bits of information that combine across large neural populations to encode shape, color, texture, and other visual characteristics. Measuring these signals requires invasive microelectrode recording, which in humans is limited to rare clinical situations (Quiroga et al. 2005). Thus, systematic study of visual coding dimensions is mainly limited to non-human primates, especially macaque monkeys, which have visual perceptual capacities comparable to those of humans. Here, I will review our recent studies of visual shape coding in the monkey ventral pathway and speculate on their implications for understanding depictive space.

Perception depends on transformation of visual dimensions

At early visual processing stages, the coding dimensionality is dictated by the input pattern on the retina, which is a 2D, isomorphic representation of the light image formed in the eye. This isomorphic representation is distributed across approximately 10^6 retinal ganglion cells, which function like pixels to signal brightness or color contrast in a local image region. Thus, the coding dimensions at this stage are the two spatial dimensions of the visual image, and each neuron represents contrast at some point in the 2D space they define.

This early, pixellated, isomorphic representation is not a suitable ground for visual cognition. It is highly distributed across a huge number retinal ganglion cells, making the signal too large and complex for rapid, efficient access by other brain regions. It is highly inconsistent – the retinal image of a given object changes drastically from moment to moment due to constantly varying position, distance, orientation, lighting, occlusion, and in some cases plastic deformation – making it impractical for recognition or memory storage. Finally, the retina explicitly represents only local contrast, not useful, meaningful information about objects in the environment. We cannot comprehend our world in terms of dot patterns; a pointillist painting makes no sense at close range. Visual experience depends on explicit neural signals for larger-scale constructs.

To make visual information useful, and to make visual thought possible, the ventral pathway must transform the distributed, inconsistent retinal image into a representation that is much more compact, consistent, and explicit about real-world object characteristics. This transformation yields stable information that is decipherable to other brain regions, storable in memory, and reportable in speech. How or why conscious visual experience emerges from this object-related, compact, consistent representation is a question at the nexus between neuroscience and philosophy. Whether and how that question will be answered is intensely debated. But however such metaphysical issues are resolved, neuroscience can provide a fascinating insight into the structure of the representation – the multi-dimensional space that defines the nature and extent of visual thought.

Contour orientation

The first major step in the dimensionality transformation leading to visual cognition is the extraction of local orientation in V1 (primary visual cortex). In V1, representation of shape is still largely spatiotopic – neurons have small receptive fields encoding fixed retinal locations that together tile the visual image. However, V1 receptive fields are larger, and thus position coding is less precise. The substitute for that loss in positional precision is information about local contour

orientation (Hubel & Wiesel 1968). Locally elongated contrast regions are represented not as a large number of constituent pixels, but in the same way as a line in Cartesian geometry, in terms of a single position value and an orientation or slope, i.e. the derivative of one spatial dimension with respect to the other. Thus, the representation of shape becomes more compact by taking advantage of a ubiquitous characteristic in visual images – the presence of locally extended contours. The representation also becomes slightly more consistent, in that some V1 neurons respond to the same orientation across a (small) range of positions. Finally, because contour edges contain most of the shape information in our world, orientation extraction is a step toward making object-relevant information more explicit.

The V1 transformation is only a first step toward the compactness, consistency, and explicitness required for visual cognition. The V1 representation of shape is still highly distributed across a spatiotopic neural map, extremely variable across minor image changes, and contains no explicit signals for spatial constructs of the size and complexity we experience. The major work of transformation into a relevant, compact, consistent dimensionality remains to be done at further stages in the ventral visual pathway.

Higher-level dimensions of visual coding

The nature of this further transformation is largely unknown. However, our recent results shed some light on how shape coding dimensions progress beyond local orientation. We find that a higher-level derivative, curvature, is extracted at the V4 level, enabling explicit, compact representation of larger contour fragments (Pasupathy & Connor 2001; Pasupathy & Connor 2002; Pasupathy & Connor 1999). Sensitivity to local curvature is even more prevalent at subsequent processing stages in inferotemporal cortex (IT) (Brincat & Connor 2004), which comprises cortical region proceeding in a posterior to anterior direction, with probable homology to human areas V8 and LOC. In parallel to the emergence of curvature tuning, retinotopic position tuning is gradually transformed into object-relative position tuning, thus achieving consistency against translation. The evidence for these dimensionality changes is reviewed in this section.

In our world, visual image contrast information can be summarized at a local level in terms of orientation, i.e. the axis of contour elongation. In V1 and V2, where receptive fields are small, orientation is the major dimension for coding shape information. At progressively higher levels in the ventral visual pathway, neurons have larger receptive fields and represent larger image regions. In area V4, receptive fields scale roughly with eccentricity about the fovea; in other words, at a point 5 degrees of visual angle away from the center of gaze, V4 receptive fields are approximately 5 degrees in diameter. As a result, V4 receptive fields encompass

significant portions of any given object (or entire small objects). In posterior IT, receptive field size is on the order of twice that in V4, and in anterior IT receptive fields can span most of the visual image.

On these larger scales, contours cannot be summarized with a single orientation value, because orientation changes along the length of most contours. However, the brain has discovered a higher derivative, curvature, that enables compact description of larger contour fragments. Curvature is the rate of change in orientation per unit contour length. In our world, due to physics, material properties, biological structure and artifactual construction, many contours have fairly constant curvature over regions on the order of V4 receptive field sizes. Not surprisingly, the brain takes advantage of this statistical consistency – contour curvature is one of the major shape-related tuning dimensions in area V4 and at higher processing stages.

We have demonstrated neural tuning in the curvature dimension by studying the shape responses of individual neurons in areas V4 and posterior IT of macaque monkeys performing a visual fixation task (Brincat & Connor 2004; Pasupathy & Connor 1999; Pasupathy & Connor 2001; Pasupathy & Connor 2002). Stabilizing eye position is necessary in order to present visual stimuli consist within the neuron's receptive field. The fixation spot and shape stimuli are presented on a video screen under computer control. We parametrically varied the curvature, orientation, and object-relative position of intermediate-sized contour fragments, combining multiple fragments in a factorial cross to create stimuli with various configurations of convex protrusions and concave indentations. Depending on the range of contour fragments, this produced sets of 350–600 shapes. Typical neurons responded to a broad subset of these shapes with action potential firing rates ranging from near 0 to around 25–50 spikes per second.

Visual inspection of these response patterns revealed that the most effective stimuli, while differing in global shape, were characterized by consistent local shape properties. For example, a given neuron might respond strongly to all shapes containing a sharp convex projection pointing upwards and positioned to the right of object center. Another neuron might respond strongly to all shapes containing downward concave curvature below object center. Response levels varied gradually with moderate shape changes – a neuron responding at 30 spikes per second to shapes containing sharp convex curvature near the top might respond at 15 spikes per second to broader convex curvature near the top or sharp convex curvature near the upper right or upper left.

These graduated response characteristics were quantifiable with Gaussian functions in much the same way as orientation tuning in V1. We have used up to 6 dimensions to quantify local contour geometry: curvature, orientation (the angle of straight edges, or the direction in which curves “point”), x and y object-relative position, x and y absolute position. Any given shape comprises multiple contour

fragments corresponding to multiple points in contour geometry space. Most V4 neurons can be well-characterized with a single Gaussian tuning function. In other words, their response to any given shape is predicted by how closely the contour fragment points defining that shape approach the peak of the Gaussian. Thus, most V4 neurons signal the presence of one particular type of contour fragment, in the context of many different global shapes. A reasonable hypothesis following from this observation is that, at the V4 level, entire shapes are represented by ensembles of neurons signaling the presence of their constituent parts. This hypothesis is examined in the following section.

Population coding in terms of object parts

To analyze population coding in area V4, we combined information across experiments for 109 neurons to estimate the population response patterns evoked by individual stimuli (Pasupathy & Connor 2002). This analysis is based on the vector averaging approach developed by Kenneth Johnson (Mountcastle 1995) and first implemented in Georgopoulos' studies of reach direction coding in motor cortex (Georgopoulos et al. 1983). In effect, each neuron is considered to "vote" for the presence of a contour fragment corresponding to its tuning peak. The strength of that vote at any point in time is the neuron's response rate. For a given stimulus, the votes (response rates) are averaged across neurons and smoothed to produce a response surface in curvature/orientation/position space.

We found that this response surface contained peaks corresponding to the major contour features of the stimulus in question. For example, the V4 population response to a teardrop shape comprised a peak corresponding to very sharp (angle-like) convex curvature above object-center, peaks corresponding to concave curvature to the upper right and upper left, and a peak corresponding to broad convex curvature near the bottom. For each stimulus that we studied, the major constituent contour features were represented in this way. Moreover, the original stimulus shape was derivable from the neural population response alone. This "decoding" was accomplished with an algorithm that found local peaks in the population response and translated these into control points for a cubic spline function used to draw the estimated stimulus boundary. It is doubtful that the brain decodes its own representations in this way, but the reconstruction exercise demonstrates that the multi-peaked population representation of boundary shape in area V4 is complete and relatively explicit, in the sense of being straightforward to decode.

Integration of object part information

Beyond V4, in the posterior regions of IT cortex, neurons have larger receptive fields and more complex response properties. We found that, as in V4, these complex shape sensitivities could be quantified with tuning functions in a multi-dimensional space based on local contour geometry (Brincat & Connor 2004). However, whereas in V4 a single tuning peak captured most of the explainable response variance, in posterior IT 2–4 peaks were typically required. In other words, IT neurons are sensitive to multiple types of contour fragments, as though they are synthesizing inputs from multiple V4 neurons. As in V4, this sensitivity is defined by local contour geometry and relatively independent of global shape. Thus, a given IT neuron responds strongly to a wide variety of stimulus shapes, all sharing sets or subsets of characteristic contour features. In many cases, neurons are also inhibited by one or more ranges of contour features.

For some IT neurons, this integrative response property is mainly linear or additive. For example, a given neuron might respond at 10 spikes per second to shapes containing a concavity at the lower right, 15 spikes per second to shapes containing a concavity at the upper left, and 25 spikes per second to shapes containing both concavities. By themselves, these neurons provide somewhat ambiguous information, because any given response rate can correspond to a variety of combinations of contour features. The veridical combination of contour features would still be decodable at the population level.

Other neurons integrate contour fragment signals in a more nonlinear fashion. For these neurons, responses to shapes containing the individual contour fragments tend to be low, but responses to shapes containing the appropriate configuration of multiple fragments are high. These selective responses provide much more explicit information about complex shape configurations. A given response rate, especially near the maximum for the neuron in question, corresponds to a limited range of similar configurations. The representation carried by these nonlinear neurons is also more compact – it takes fewer neurons to encode a given shape. Finally, specific configurations of contour fragments are highly relevant to recognition and discrimination of the types of objects encountered in our world, which frequently contain characteristic part shapes. Thus, the integration of information across multiple contour regions in IT creates a representation that is more compact, explicit, and relevant.

Our most recent analyses show how these linear and nonlinear integration processes evolve through time (Brincat & Connor 2006). Neurons with primarily linear responses tend to have earlier response onsets (beginning around 70 milliseconds following stimulus onset) and their response rates peak early. Neurons with primarily nonlinear responses (i.e. neurons selective for multi-fragment configurations) tend to have later onsets and later peak responses. Many neurons

transition from more linear responses at early time points to more nonlinear responses at later time points. Across the entire neural population, linear signals peak at around 120 milliseconds after stimulus onset, while nonlinear signals peak about 60 milliseconds later, at around 180 milliseconds after stimulus onset. This reflects a dynamic process of part synthesis leading toward the more compact and explicit representation carried by nonlinear responses to part configurations.

Relationship of higher-level neural representation to phenomenal visual experience

The structure of phenomenal visual experience is intimately related to and in some sense dependent upon the dimensionality of neural representations. It could be argued that neural tuning dimensions completely define and constrain the domain of visual thought. I believe some of the results described above begin to illuminate the relationship between phenomenal structure and neural dimensionality.

First, the neural tuning dimensions we observed at higher stages in the ventral pathway are strikingly consistent with visual experience. Curvature is strongly represented by neurons beginning at the V4 level, and curvature is a ubiquitous and fundamental aspect of shape experience, for which we have a rich vocabulary used extensively in describing objects. Curvature presents itself as an immediate and primary shape attribute (Andrews et al. 1973; Treisman & Gormican 1988; Wilson et al. 1997; Wolfe et al. 1992), with no introspectively observable process of derivation, even though a computationally substantial transformation of the retinal image must be implicitly performed to extract such a higher-level derivative. The overwhelming phenomenal presence of curvature is arguably due to the large number of neurons with strong tuning for curvature at higher levels in the ventral visual pathway. What the brain makes most explicit in neural population responses has the greatest phenomenal potency. Likewise orientation, a lower derivative, is highly explicit at the neural and phenomenal levels. At a higher order, the derivative of curvature may also be represented neurally (since some V4 cells appear to be selective for spirals (Gallant et al. 1993; Gallant et al. 1996)). This derivative is also perceptually available (since for example we can immediately distinguish the shape of a cow horn), although our vocabulary for describing it ("curvature derivative") is limited and not in common use. Still higher order derivatives are not immediately perceivable and are probably not represented explicitly at the neural level. The brain has developed no method for extracting these higher order derivatives because they do not help in characterizing the shape elements found in our world.

Likewise, the neural transformation from retinotopic position tuning toward object-relative position tuning is reflected by our phenomenal experience and perceptual reports. We readily describe relative position within local reference frames

defined by objects (e.g. “the pen is about two inches to the right of the book”), but we have much less precise access to absolute position. This seems trivial and obvious, but only because our neural representation of pictorial space is structured in terms of local reference frames, and this determines how we think about visual space. Higher-level neurons are tuned precisely for relative position but generalize broadly across absolute position, and this gives us our precise sense of relative position. It seems ridiculous to imagine reporting the positions of objects in terms of exact Cartesian coordinates (at least without measuring devices), yet consider that all image information is originally encoded in an absolute (retinotopic) reference frame, at very fine resolution (on the order of retinal ganglion receptive field sizes). The computationally easiest thing for the brain would be to continue to operate in retinotopic coordinates. Instead, because relative position is so much more relevant for operations in our world, the brain performs a difficult transformation into flexible local reference frames, and this transformation defines the spatial structure of our visual experience.

Next, representation in terms of parts, a popular theory (Biederman 1987; Marr & Nishihara 1978), is strongly confirmed by our population analysis of V4 and the continuation of parts-level coding we observed in IT. Correspondingly, perceptual and cognitive segregation into parts is fundamental to our experience and knowledge of objects. When we are asked to describe an object, the description is typically in terms of its parts and where they are in relation to each other (the object-relative position coding dimension). Again, this seems trivial only because it is so completely determined by the neural structure of our visual representations that it is hard to imagine anything else. But a very different and wholly conceivable possibility is the type of holistic coding scheme imagined by Shimon Edelman – the “Chorus of Prototypes” model in which individual neurons signal not parts but the overall shape similarity of a given stimulus to previously learned prototypes (Edelman 1999). If our brain implemented such a scheme, we would not be able to report object parts and their relationships; instead we would describe shapes as mixtures of familiar objects. In fact, highly learned expert within-category shape discriminations like face recognition probably do depend on holistic coding. That is why we describe faces as “feminine”, “Asian”, or “like Humphrey Bogart”, rather than in terms of precise distances between the eyes, nose and mouth. For generic object perception, however, our experience and our language is structured in terms of prototypical parts. If it were not, we would in a sense be unable to “see” a completely novel shape with no similarity to familiar objects; it would appear as a nonsensical jumble. In reality, there is no fanciful object I could sketch that you would be unable to appreciate, describe, and remember, because of the combinatorial power of the parts-based coding scheme adopted by the visual system.

Finally, the temporal evolution of the neural representation following stimulus onset has clear relevance to understanding how visual perception progresses on a millisecond time scale. Basic object categorization can occur very rapidly (Thorpe et al. 1996), in a way that does not seem to depend on complex configurational processing (Rousset et al. 2003). Rapidly emerging neural signals for individual parts could support such non-configural categorization. In contrast, fine discriminations based on multi-part configurations require more time (on the order of 50 ms) (Arguin & Saumier 2000; Wolfe & Bennett 1997). The gradual emergence of neural signals for multi-part configurations could explain this delay in refined shape appreciation.

The neural/phenomenal relationships suggested above are speculative. They could be supported experimentally by quantitative comparisons between psychophysical reports of shape experiences and neural tuning distributions (preferably within the same species), but no such evidence exists at this time. Nevertheless, these speculations exemplify how neural coding studies might inform understanding of pictorial space at the phenomenal and psychophysical levels. Psychophysical/neurophysiological comparison has been one of the most powerful tools for understanding the basis of simpler perceptual domains like motion (Newsome et al. 1989). The surprising correspondences across levels that have emerged from these studies have changed scientific thinking about how the brain gives rise to simple visual perceptions. There is no reason that the same parallel approach could not revolutionize our understanding of how complex pictorial perception arises.

References

- Andrews, D. P., Butcher, A. K., & Buckley, B. R. (1973). Acuities for spatial arrangement in line figures: human and ideal observers compared. *Vision Research*, 13, 599–620.
- Arguin, M., & Saumier, D. (2000). Conjunction and linear non-separability effects in visual shape encoding. *Vision Research*, 40, 3099–3115.
- Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94, 115–147.
- Brincat, S. L., & Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience*, 7, 880–886.
- Brincat, S. L., & Connor, C. E. (2006). Dynamic shape synthesis in posterior inferotemporal cortex. *Neuron*, 49, 17–24.
- Edelman, S. (1999). *Representation and Recognition in Vision*. Cambridge, MA: MIT Press.
- Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: recognition, navigation, or encoding? *Neuron*, 23, 115–125.
- Gallant, J. L., Braun, J., & Van Essen, D. C. (1993). Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science*, 259, 100–103.

- 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.
- Georgopoulos, A. P., Caminiti, R., Kalaska, J. F., & Massey, J. T. (1983). Spatial coding of movement: A hypothesis concerning the coding of movement direction by motor cortical populations. *Experimental Brain Research Supp.*, 7, 327–336.
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology (Lond)*, 195, 215–243.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997a). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17, 4302–4311.
- Kanwisher, N., Woods, R., Iacoboni, M., & Mazziotta, J. (1997b). A Locus in Human Extrastriate Cortex for Visual Shape Analysis. *Journal of Cognitive Neuroscience*, 9, 133–142.
- Kourtzi, Z., & Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *Journal of Neuroscience*, 20, 3310–3318.
- Lerner, Y., Hendler, T., Ben Bashat, D., Harel, M., & Malach, R. (2001). A hierarchical axis of object processing stages in the human visual cortex. *Cerebral Cortex*, 11, 287–297.
- Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society B: Biological Sciences*, 200, 269–294.
- Mountcastle, V. B. (1995). The parietal system and some higher brain functions. *Cereb. Cortex*, 5, 377–390.
- Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision 12. *Nature*, 341, 52–54.
- Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of Neurophysiology*, 82, 2490–2502. <http://www.jneurosci.org>
- Pasupathy, A., & Connor, C. E. (2001). Shape representation in area V4: Position-specific tuning for boundary conformation. *Journal of Neurophysiology*, 86, 2505–2519.
- Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience*, 5, 1332–1338.
- Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 433. Ref Type: Generic.
- Rousselet, G. A., Mace, M. J., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision*, 3, 440–455.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381, 520–522.
- Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95, 15–48.
- Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. G. Ingle, M. A. Goodale, & R. J. Q. Mansfield (Eds.), *In Analysis of Visual Behavior* (pp. 549–586). Cambridge, MA: MIT Press.
- Wilson, H. R., Wilkinson, F., & Asaad, W. (1997). Concentric orientation summation in human form vision. *Vision Research*, 37, 2325–2330.
- Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research*, 37, 25–43.
- Wolfe, J. M., Yee, A., & Friedman-Hill, S. R. (1992). Curvature is a basic feature for visual search tasks. *Perception*, 21, 465–480.

Boundary Gestalt limits flow Gestalt

The geometry of good continuation

Ohad Ben-Shahar and Steven W. Zucker

Introduction

Orientation enjoys a distinguished position among the Gestalt laws of good continuation, since it is the basis for many of the classical demonstrations. However these were essentially one-dimensional, in the sense that they involved good continuation of curves through (non-simple) crossing points rather than imposing discontinuities (Fig. 1(top)). However, many visual structures, such as optical flow [Gibson 1950] and oriented texture [Kass, Witkin 1987; Ben-Shahar, Zucker 2003] also appear as flow fields over the image plane, but these are two-dimensional in the sense that they cover regions of an image. Understanding the geometry of these flows is a key step toward the interpretation of visual scenes, both two dimensional [Kanizsa 1979; Kass, Witkin 1987] and three dimensional [Stevens 1983; Todd, Reichel 1990]. In this paper we argue that similar techniques are relevant for understanding both the classical (1-D) Gestalt examples and these 2D flows.

Perceptually, visual flow fields are characterized by their dense, smoothly varying (almost everywhere) oriented structure. This is often related to surface properties, and our visual systems may compromise the veridicality and the accuracy of a description of a flow to obtain one that conveys a consistent, well-defined surface. As shown in Figure 1(bottom), for example, the different numbers of parallel surface contours are integrated into a coherent percept, with good continuation apparent along them. But this result is mathematically incorrect; the discontinuities, which are behind the occluders, are not immediately apparent. Scrutiny, or counting the lines, is necessary to detect the differences. Perceptually, then, some very powerful inferring is taking place.

Abstractly a flow is an orientation function $\theta(x, y)$ defined over the image plane. In this paper we shall review several different instances of the orientation function, illustrating how boundaries [Parent, Zucker 1989], oriented textures [Ben-Shahar, Zucker 2003], shading [Breton, Zucker 1996], and (by implication)

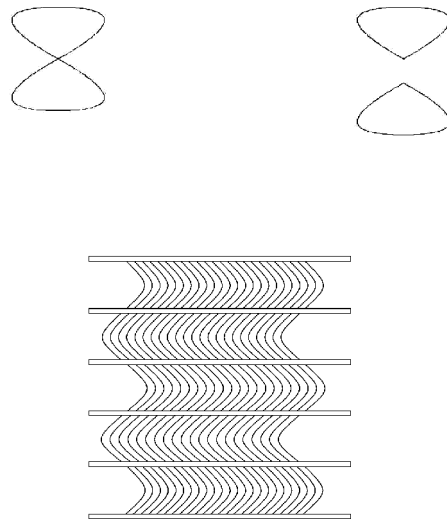


Figure 1. Examples of Gestalt good continuation in orientation. (top) The classical figure “8” display is seen as a single continuous curve, rather than the non-generic conjunction of two special ones. (bottom) Good continuation for oriented texture flows, however, offers additional challenges. Notice, in particular, whether there is a constant number of contours comprising this texture between each of the occluders.

color [Ben-Shahar, Zucker 2004] (in particular, hue) can all be treated as flows over regions of the image plane.

To support inferences about surfaces, it is essential to realize that these different flows are not independent from one another. For example, the relationship between the geometry of shading and that of its bounding edge provides a basis for classifying edges [Breton, Zucker 1996], and can be used to resolve occlusion relationships [Huggins, Zucker 2001]. Since shading flow boundaries (i.e., the curves along which the flow should be segmented into coherent parts) are defined as intensity edges, shading is a clear example of a visual flow for which both the flow field and its boundaries can be directly measured from the image. Moreover, hue flows, as we have shown [Ben-Shahar, Zucker 2004], enjoy somewhat different relationships with boundaries. Taken together with shading, they provide a very rich source of information about the physical scene. Thus in this paper we focus on the cases of boundaries, texture, and shading.

All of these flows suffer the same basic problems with regard to measurement: If the initial measurements of a flow field are made locally, with sensors like V1 receptive fields, they are likely to be imperfect, suffering from quantization and noise in some regions and perhaps failing altogether in others. Hence, an effective computational process for the inference of coherent visual flow must be able to do so from a noisy, incomplete set of measurements. Furthermore, as implied from

the discussion above, it should localize singularities, reject non-flow structures, and behave appropriately along line discontinuities and boundaries.

We have developed a computational model that addresses all these issues within a framework that is inspired by the columnar architecture of the primary visual cortex [Ben-Shahar, Zucker 2001; Ben-Shahar, Zucker 2003]. In this chapter we review that framework, illustrate the computation of different types of flows within it, and illustrate some of the interactions between them. All are consistent with the architecture of visual cortex, and all suggest that visual cortex embodies computations that extend a good distance toward the qualitative description of surface properties.

We begin with a very short review of the neural architecture of the first visual area, to set the basic outlines of our framework. We then develop the general structure of a flow in terms of the relevant differential geometry, and observe how this fits naturally into the neural architecture. With this background in place, we then illustrate the boundary, texture, and shading flows. Finally, we show how to integrate boundary information directly into the flow computation.

A cortical columnar machine for early vision

Visual cortex is organized to a significant extent for the processing of orientation information; that is, around selective responses to local oriented bars. In a classical observation [Hubel, Wiesel 1977], recordings along a tangential penetration encounter a sequence of cells with regular shifts in orientation preference, while normal penetrations reveal cells with similar orientation and position preferences but different receptive field sizes. Together they define an array of orientation columns which, combined with eye of origin, provide a representation for visual information processing. Together with network dynamics, these columns represent an instance of what might be described as a visual cortical columnar machine specialized for problems in early vision. Although significant feedforward and feedback projections exist and may influence the processes we describe, we shall not discuss them further in this chapter. Our mathematical position in effect allows us to abstract over articulations beyond V1.

The (position, orientation) representation suggests two key abstractions for early vision: (i) local estimates of orientation are modified via relationships with estimated orientations at nearby positions; and (ii) multiple orientations along a single column can signal an orientation discontinuity (Figure 3). But there are limits; clearly all orientations at every position should not be allowed. Thus constraints are needed for flows over neighborhoods, and the mathematical analysis of visual flows provides them.

Given that the initial measurements of a visual flow field may contain spurious or missing values, we would like to refine the flow field to counteract these effects. Interpolating and fitting [Rao, Jain 1992], smoothing [Perona 1998], and diffusing [Tang, Sapiro, Caselles 2000] the orientation function $\theta(x, y)$ corresponding to the flow are commonly used approaches to achieving this goal, but they are prone to affect the underlying geometry of the flow in undesirable ways. In particular, they can distort flow singularities that must be preserved to correctly interpret visual scenes [Ben-Shahar, Zucker 2001; Ben-Shahar, Zucker 2003]. To overcome this problem we process the visual flow by enforcing local coherence; that is, by ensuring that each local measurement of the flow field is consistent with its neighboring measurements.

The differential geometry of curves and flows

To motivate our approach, and to introduce the abstractions of tangent, curvature, and frame, we first examine the differential geometry of visual curves; and then proceed to the geometry of flows. Observe, in both cases, that any local measurement of orientation by, say, a simple or a complex cell, provides an estimate of the best linear approximation within a neighborhood defined by the spatial support of the cell's receptive field. This, in mathematical terms, is an estimate to the tangent to the curve (provided the curve is smooth and the tangent exists).

The differential geometry of curves

The problem of curve detection, within the framework of Figure 2, is that local estimates of the tangent orientation are noisy; the natural solution is to exploit the contextual relationships between nearby tangents. We begin with the observation that local measurements of the tangent to a curve at a point and at a nearby point in its neighborhood will usually have different orientations. Therefore one cannot simply extend the tangent in the direction it is pointing; so-called, colinear facilitation. But there is a functional relationship between them, and it is given by the Frenet notion of transport [O'Neill 1966]: if the neighboring measurement of orientation is transported along the curve to the original point, then it should agree with the orientation measured at that point if they are consistent. If they disagree, then they are not consistent; see Figure 4.

Unfortunately, however, the curve is unknown, so transport is impossible in the above formulation. However, if the curvature could also be measured [Dobbins, Zucker, Cynander 1987] at the point, then the transport could be along an approximation to the curve (its osculating circle) instead. Since the transport is

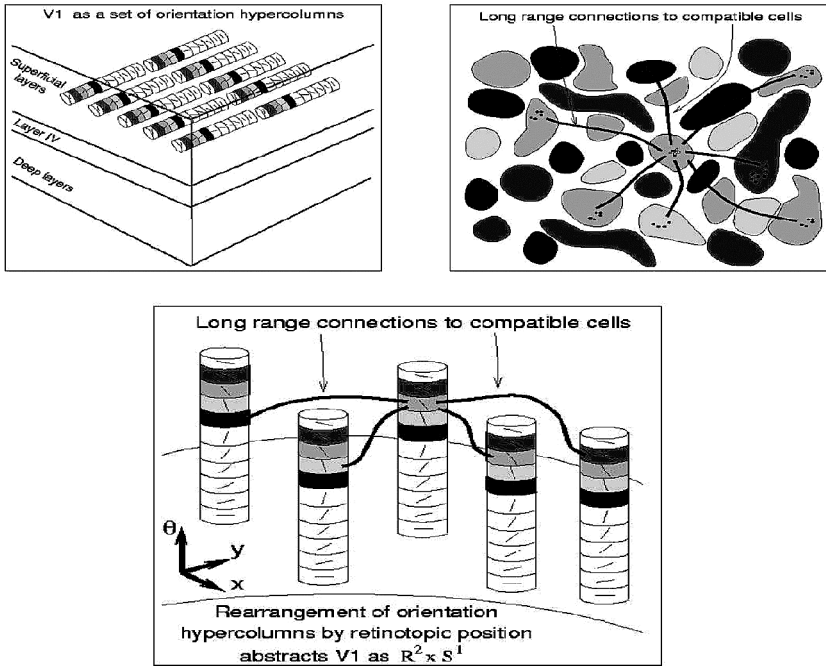


Figure 2. Abstracting the functional architecture of visual cortex. (top, left) The standard Hubel-Wiesel “ice cube” model, which, although it is a cartoon, expresses the fundamental observation that each local retinotopic area is covered by receptive fields that span a range of orientations (several tangential penetrations are shown, each of which corresponds to cells with spatially overlapping receptive fields but tuned to a different orientations from the same eye of origin). (top, right) “Flattened” cartoon of the axon from one cell, showing how it forms long-range horizontal connections with cells selective for some orientation at nearby positions. The irregular blobs denote orientation domains [7]. (bottom) A re-drawing of the ice-cube model, emphasizing a retinotopic array (the tilted plane in (x,y) coordinates) and columns of superficial-layer cells (re-drawn from (top, left)) but now including the long-range horizontal connections. Our goal is to develop a model for these connections.

only over small distances, then the error incurred by this operation is small. This is the essence of differential geometry, and this is the formal notion of co-circularity; see Figure 4 and [Parent, Zucker 1989].

The differential geometry of flows

A natural representation of a visual flow which highlights its intrinsic geometry is its *frame field* [O’Neill 1966]. Here a local frame $\{E_T; E_N\}$ is attached to each

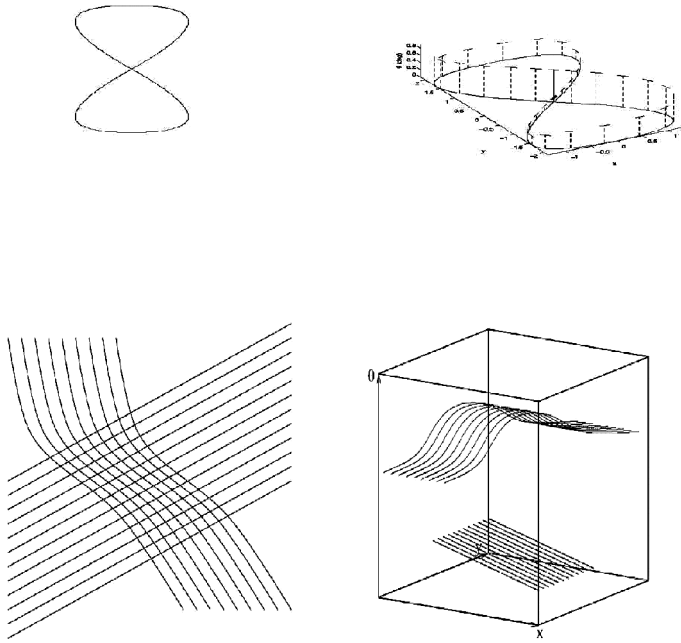


Figure 3. Multiple orientation-valued functions in position separate when lifted to (position, orientation) space. (top) The non-simple figure “8” curve in the image becomes a simple curve when lifted into (position, orientation) space, since now the different orientations at the same point are separated. (bottom) Overlapping textures separate into smooth flows at different “layers” of orientation.

point q of the flow, with E_T tangent and E_N normal to the flow. (Notice how this generalizes the case of curves discussed above.) Small translations in direction V from the point q rotate the frame, a change which is characterized through the covariant derivatives $\nabla_V E_T$ and $\nabla_V E_N$ of the underlying pattern. Cartan’s connection equation [O’Neill 1966] expresses these covariant derivatives in terms of the frame field itself:

$$\begin{pmatrix} \nabla_V E_T \\ \nabla_V E_N \end{pmatrix} = \begin{bmatrix} 0 & w_{12}(V) \\ -w_{12}(V) & 0 \end{bmatrix} \begin{pmatrix} E_T \\ E_N \end{pmatrix}. \quad (1)$$

The connection form $w_{12}(V)$ is a linear function of the tangent vector V and can thus be represented by two scalars at each point. In the basis $\{E_T, E_N\}$ these scalars are defined as $K_T = w_{12}(E_T)$ and $K_N = w_{12}(E_N)$, which we call the *tangential curvature* and the *normal curvature* – they represent the rate of change of the flow’s dominant orientation in the tangential and normal directions, respectively. In terms of (x, y) and its differential, these curvatures become:

$$\begin{aligned} K_T &= d\theta(E_T) = \nabla\theta \cdot E_T = \nabla\theta \cdot (\cos\theta, \sin\theta) \\ K_N &= d\theta(E_N) = \nabla\theta \cdot E_N = \nabla\theta \cdot (-\sin\theta, \cos\theta) \end{aligned} \quad (2)$$

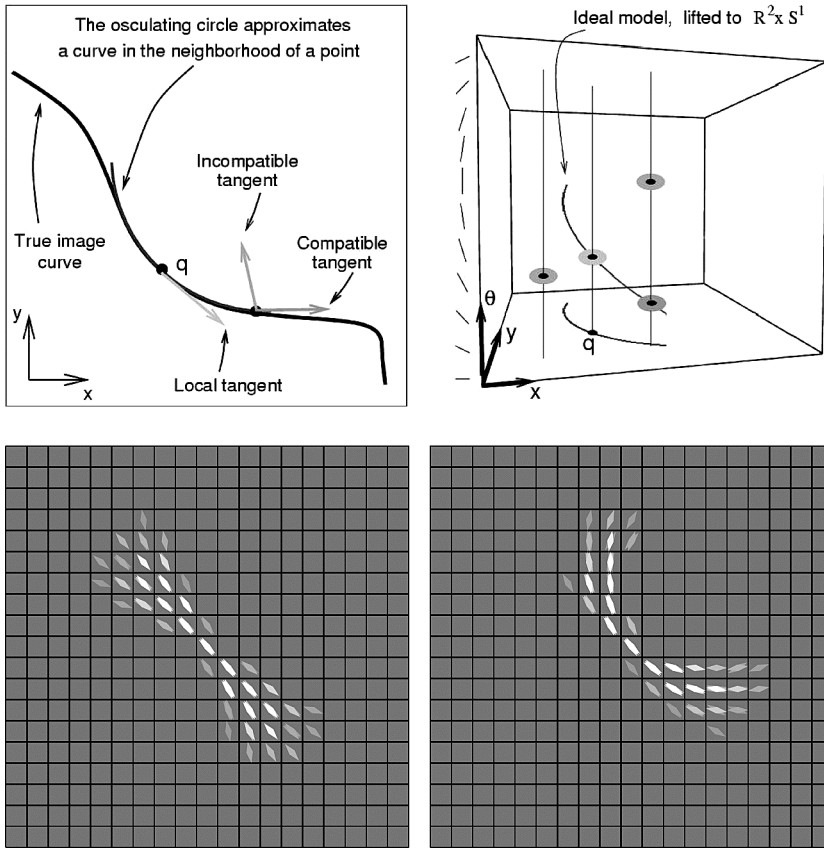


Figure 4. The differential geometry for curves. (top) The geometry displayed in image coordinates (left) and lifted to (position, orientation)-space (right). Notice how the local approximation of the curve by its osculating circle is close to the curve, so that nearby tangents can be transported along it without incurring much error. (bottom) Discretization of the connections projected onto image space for curvature straight (left) and curved (right). It is these connections that define the compatibility between two nearby tangents; they are clearly a function of curvature. Amount of support is indicated by brightness; only positive values shown.

Knowledge of E_T , E_N , K_T , and K_N at a point q enables us to construct a local approximation to the flow which has the same orientation and curvatures at q ; we call such an approximation an *osculating flow field*. The osculating flow field is important in that it predicts flow values in the neighborhood of q in direct analogy to the way that the osculating circle provided a prediction for co-circularity over curves. Comparing these predictions to the measured flow values indicates how consistent the measured values of the flow at q are with those at its neighbors and suggests how to update them to be consistent.

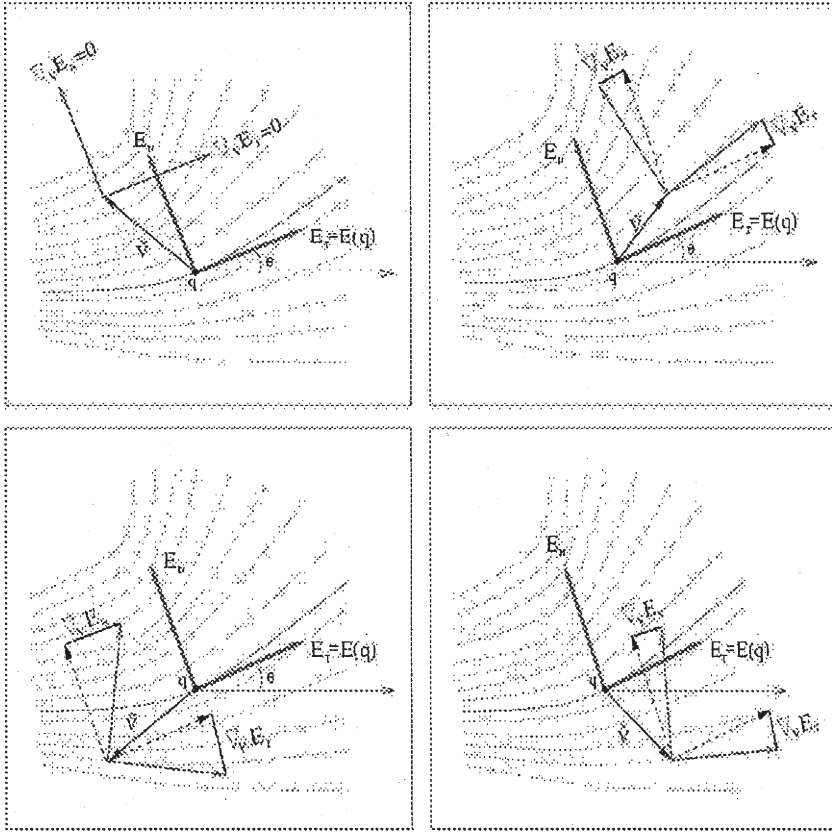


Figure 5. The frame field representation of visual flows. The local behavior of the frame is described by its covariant derivatives $\nabla_V E_T$ and $\nabla_V E_N$ which are always normal to E_T and E_N , respectively. Since the connection form – the operator that describes the frame’s rotation for any direction V – is linear, it is fully characterized by two numbers computed as projections on two independent directions. In the basis of the frame this yields the curvatures κ_T and κ_N .

There are an infinite number of possible osculating flow fields to choose from. However, there exist criteria for “good” osculating flow fields. One such criterion is the minimization of the harmonic energy $E\theta = \int \int \|\nabla\theta\|^2 dx dy$ associated with the orientation function of the flow, as is used in orientation diffusion [Tang, Shapiro, Caselles 2000]. Viewing the orientation function as a surface in $R^2 \times S^1$, however, suggests that the osculating flow field should minimize the surface area $A[\theta] = \int \int \sqrt{1 + \theta_x^2 + \theta_y^2} dx dy$. Finally, the duality of the curvatures K_T and K_N suggests that the osculating flow field should exhibit unbiased curvature covariation.

Surprisingly, there is a unique osculating flow field which satisfies all of these criteria simultaneously [Ben-Shahar, Zucker 2001; Ben-Shahar, Zucker 2003]. In

the space $R^2 \times S^1$ of orientations over the image plane it takes the form of a right helicoid; this is the generalization of the osculating circle that we use.

Armed with a model of the local structure of visual flow we are in a position to compute a globally coherent flow, the procedure for which is described in the next section.

Computing coherent visual flows

The advantage of having a model for the local behavior of visual flow lies in the ability to assess the degree to which a particular measurement is consistent with the context in which it is embedded. This, in turn, can be used to refine noisy measurements, remove spurious ones, and fill in “holes” so that global structures become coherent. A framework in which one can pursue this task by iteratively maximizing the average local consistency over the domain of interest is *relaxation labeling* [Hummel, Zucker 1983]. We have developed such a network for the organization of coherent visual flows [Ben-Shahar, Zucker 2003]. The following is a short overview of that system.

A direct abstraction of the relaxation process for visual flow should involve an image-based 2D network of nodes $i = (x, y)$ (i.e., pixels) whose labels are drawn from the set $\Lambda = \{\text{no-flow}\} \cup \{(\theta, K_T, K_N) | \theta \in (-\pi/2, \pi/2), K_T, K_N \in [-K, K]\}$ after the appropriate quantization. To allow for the representation of either “no-flow” or multiple flows at a pixel, we replace this abstraction with a 5D network of nodes $i = (x, y, \theta, K_T, K_N)$ whose labels are either TRUE (T) or FALSE (F). For each node i , $p_i(T)$ denotes the confidence that a visual flow of orientation θ and curvatures K_T, K_N passes through pixel (x, y) . Since $p_i(F) = 1 - p_i(T)$ we need to maintain and update the confidence of only one label at each node.

The geometrical compatibilities $r_{ij}(\lambda, \lambda')$ that drive our relaxation process are computed from the osculating flow field as defined by the right helicoid. Measurement quantization implies that every possible node i represents an equivalence class of measurements, each of which induces a field of compatible labels in the neighborhood of i . In the continuum, the union of all these fields forms a consistent 5D “volume” that after quantization results in a set of excitatory labels.

With the network structure, labels, and compatibilities all designed, one can compute the support $s_i(\lambda)$, that label λ at node i gathers from its neighborhood. $s_i(\lambda)$ is typically the sum of the individual support of all labels j in the neighborhood of i

$$s_i(\lambda) = \sum_j \sum_{\lambda'} r_{ij}(\lambda, \lambda') p_j(\lambda') \quad (3)$$

Having computed the support for a label, $s_i(\lambda)$ is then used to update the confidence $p_i(\lambda)$ by gradient ascent, followed by non-linear projection. Under the 2-label paradigm and the appropriate weighting of negative (F) versus positive (T) evidence [Ben-Shahar, Zucker 2003], the projection operator takes a particularly convenient form and the update rule reduces to

$$p_i(\lambda) \leftarrow \prod_0^1 (p_i(\lambda) + \delta s_i(\lambda)) \quad (4)$$

where $\prod_0^1(x)$ projects its operand to the nearest point on the interval $[0, 1]$ and δ is the step size of the gradient descent.

While the relaxation labeling network described is an abstraction based on the differential geometry of flow fields, it is motivated by the architecture of the primary visual cortex. The columnar structure of V1 clearly lends itself to the representation of orientation fields [Hubel, Wiesel 1977], and is capable of the necessary curvature computations [Dobbins, Zucker, Cynander 1987]. Considerable speculation surrounds the functional significance of long-range horizontal connections [Gilbert 1992] between orientation columns; we posit that they play a role like the compatibility structures of our network (Fig. 6; compare with Fig. 4, bottom) [Ben-Shahar, Zucker 2003].

An example of a flow computation is shown in Figure 7. This illustrates how dense measurements can be filled-in; how noisy initial values can be corrected; and how the different orientation “layers” (more formally, sections) remain separated.

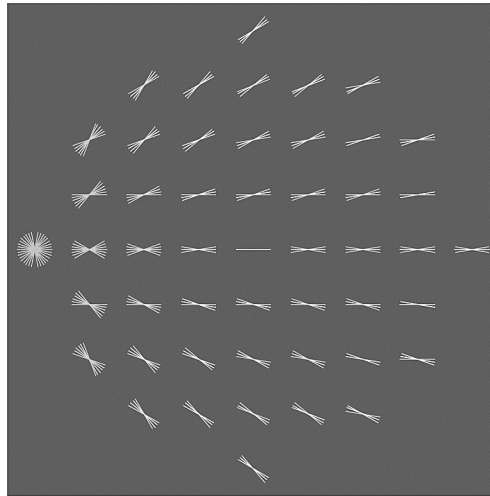


Figure 6. Example of a compatibility field for texture and shading flows with tangential curvature 0 but positive normal curvature. Note how this field introduces a singularity; it is analogous, in a local sense, to an optical flow field approaching a vanishing point.

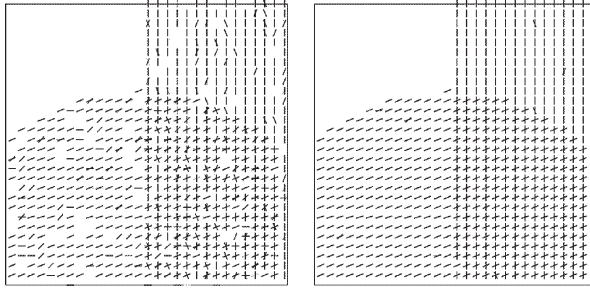


Figure 7. Illustration of the texture flow computation on an example of overlapping textures. Note how incomplete values are filled-in, how noisy values are corrected, and how the interaction between the textures at different “layers” of (position, orientation)-space is negligible.

In effect, this computation is formulated ideally for neighborhoods of interaction well within a textured region. What is not shown in this example is how values can leak across boundaries depending on their properties. We turn to this neglected point next.

Stability at discontinuities

In computing coherent visual flows it is important to respect discontinuities, as these often correspond to significant physical phenomena. The relaxation process described above does not destroy these structures because, in the high dimensional space in which it operates, the flow structures that meet along a line discontinuity, either in orientation or curvature, are separated and thus do not interact. However, without proper tuning, the relaxation process will quickly shrink or expand the flow in the neighborhood of boundaries. It is this behavior we seek to suppress. To achieve stability we normalize the compatibility function, and thus the support function $s_i(\lambda)$, to account for reduced support in the neighborhood of a discontinuity. Given the compatibility volume V_i which corresponds to a particular node i , we compute the maximal support a node can receive, $s_{\max}$, as the integral of the compatibility coefficients assuming a consistent flow traverses V_i with all supporting nodes at full confidence (Fig. 8).

It is clear that the closer i is to a flow discontinuity, the less context supports it. At the discontinuity, the flow should neither grow nor shrink, leading us to define the minimal level of support for which no change in confidence occurs, $s_{\min}$. Observe that $s_{\min}$ depends on both the geometry of the discontinuity and the minimally accepted confidence of the supporting nodes. For simplicity we assume the discontinuity (locally) occurs along a straight line. The support from neighboring

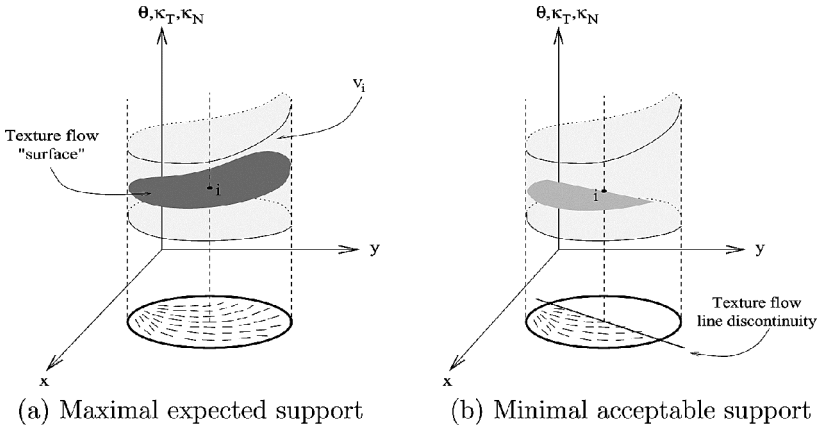


Figure 8. Practical stability of the relaxation labeling process at line discontinuities in the flow can be achieved through the normalization of the support function. (a) At each node i , $s_{\max}$ is determined by integrating the support gathered from a full confidence, compatible flow that traverses the entire compatibility volume V_i . (b) The minimal accepted support $s_{\min}$ of a flow of some minimally accepted confidence $\rho_{\min} < 1$ (depicted here by the brighter surface intensity) that terminates along a line that intersects i .

nodes of minimally accepted average confidence $\rho_{\min}$ (Fig. 8) can be approximated as $s_{\min} = (\rho_{\min} s_{\max})/2$. Normally $\rho_{\min}$ would be set to 0.5, which is the minimal confidence that cannot be disambiguated as the TRUE label. In the context of the two-label relaxation labeling paradigm and the gradient ascent update rule (Eq. 4), a decrease in the confidence of a label occurs only if $s_i < 0$. Thus, it remains to normalize the support values by mapping the interval $[s_{\min}, s_{\max}]$ to the unit interval $[0, 1]$ via the transformation $s_i = (s_i - s_{\min})/(s_{\max} - s_{\min})$ before applying the update rule.

The result of the normalized relaxation process is usually very good. Nevertheless, the fact that both the support function (Eq. 3) and the normalization are linear creates a delicate balance: while better noise resistance suggests smaller $s_{\min}$, it also implies that, at discontinuities, the flow will eventually grow uncontrollably. Some solutions to this problem are discussed in [Ben-Shahar, Zucker 2003]. However, in the case of shading flow fields, discontinuities are intensity edges and thus can be explicitly identified by edge detection. As we discuss below, this information can be directly embedded into the network to decouple the handling of discontinuities from the support normalization.

Edges as shading flow boundaries

Edges in images are important because they signify physical changes in a scene; hence the numerous efforts to detect them. The physical nature of an edge is often discernible from the appearance of the edge in the image. In particular, the relationship between the edge and the shading flow field in the neighborhood of the edge can be used to identify the physical cause of the edge. The shading flow field is defined as the unit vector field aligned with the iso-brightness contours of the image [Breton, Zucker 1996]. For example, the shading flow field is continuous across an edge caused by an abrupt albedo change but discontinuous across an edge caused by a cast shadow [Breton, Zucker 1996].

Significantly, occlusion edges can be distinguished on the basis of the shading flow field as well. At an occlusion edge of a smooth object, the edge results from the object's surface curving away from the viewer; we call this type of edge a *fold*. At a fold, the shading flow field is generically tangent to the edge due to the projective geometry of the situation (Fig. 9). On the occluded side of the edge the shading flow has an arbitrary relationship to the edge and is generically non-tangent; we call this side of the edge a *cut* [Huggins, Zucker 2001].

The ability to compute the flow field structure in the neighborhood of the edge is exactly what we are looking for to classify the edge. However, techniques that compute flow field structure without explicitly accounting for edges can destroy the relationship between the flow field and the edge and thus prevent the correct interpretation and classification of the edge. What we describe next is how we endow the connectivity structure of our relaxation labeling network with the ability to explicitly consider edge information and thus prevent the problem just mentioned. Naturally, this places some dependence on the edge detector used; however this is clearly preferable to completely ignoring the edge.



Figure 9. Illustration of shading flow in the neighborhood of an edge. When a shaded surface is viewed such that an edge appears, the shading flow field takes on different appearances depending on the nature of the edge. A fold occurs (a) when the surface bends smoothly away from the viewer (the typical occlusion case), and the shading flow field appears tangent to the edge. At a cut (b), the surface is discontinuous (or occluded), and shading flow is generally non-tangent to the edge.

Edges as nonlinear inhibition

Due to its physical nature, an edge can be thought of as dividing the shading flow field domain into distinct regions, implying that the computation of the shading flow on either side of the edge can and should be done separately. This is an intuitive but powerful argument: incorporating edges into the relaxation labeling network to regulate the growth of flow structure obviates the trade-off between high resistance to noise and strict stability along discontinuities we mentioned in Section 4. To implement this idea in the framework of relaxation labeling, what is needed is a specialized set of interactions between edge nodes and nearby shading flow nodes. These interactions would block the flow input if it comes from across the edge. With this input blocked, and so long as s_{min} is positive, the flow on one side of the edge will not extend across the edge, because the total support contributed to the other side will never exceed zero. This frees the selection of s_{min} from stability considerations and allows us to determine it solely on the basis of noise resistance and structural criteria. A cartoon illustrating these interactions appears in Figure 10. Interestingly, a nonlinear veto mechanism that is reminiscent of the one proposed here also exists in biological systems in the form of shunting inhibition [Borg-Graham, Monier, Frégnac 1998].

We have tested this adaptive relaxation labeling network on a variety of synthetic and natural images, two of which are shown in Figure 8. We used the

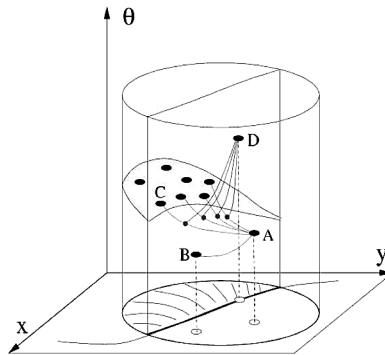


Figure 10. Edge-flow interactions for boundary stability. Assume the flow structure in the image plane is bounded by the indicated edge. Flow cell A is connected to a set of other cells (B and C) which are a part of the same coherent flow. Although A is not active (there is no flow in its corresponding retinotopic position), its facilitory interaction with the cells on the other side of the edge may eventually raise its activity level. To prevent cell C from affecting A, an active edge cell D blocks the facilitory inputs from C, thus effectively limiting A's context to cell B only. Unless enough of these cells are also active, A will not reach its activation potential, and thus will not signal any flow.

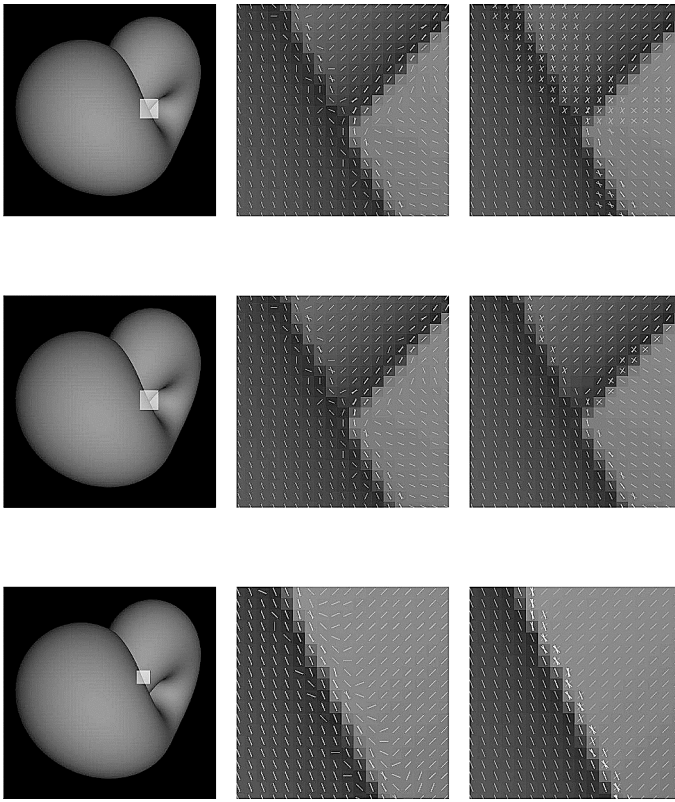


Figure 11. Examples of shading flow field relaxation without and with edges as boundary conditions. Shown (left to right): image and region-of interest (ROI); initial estimate of shading flow (thin gray segments) and edges (thick white segments) within the ROI; and the result of relaxation. (top row) Example of “T” junction without boundaries. Note how the shading flows into other areas as in the overlapping texture example. (middle row) Example of “T” junction with boundaries computed by the Canny operator. Note that even though the boundaries are incorrect at the “T” junction, the flows are correct. These flows could then be fed back to improve the edge map. (bottom row) Another example of flow with boundary along a fold. Note how the tangency condition is satisfied.

Logical/Linear [Iverson, Zucker 1995] and the Canny [Canny 1986] edge detectors and the shading flow fields were measured using standard differential operators.

Conclusions

The differential geometry of visual flows is naturally expressed in terms of Frenet Frame fields and their covariant derivatives. This implies that orientation (to ori-

ent the frames) and curvature (to specify the connection forms) are fundamental candidates for measurement. However, since any such measurements are quantized and likely to be noisy, mechanisms that enforce consistency between them are also needed. A relaxation system derived from the basic architecture of early visual cortex is readily adapted to the different types of flows, and provided the computational framework for this paper.

A key aspect of the computational framework is that boundary and flow cues can be integrated. This was illustrated in detail for the computation of coherent shading flow fields in images. It is important to capture this interaction between flows and boundaries accurately as it indicates the geometry of the scene underlying the image. Boundary information was used to adaptively alter the context that influences a given node, a mechanism which enables the network to handle noisy and sparse data sets without affecting the flow's discontinuities. Both the flow computation and the incorporation of edges as boundary conditions are readily implemented within the columnar architecture of the primary visual cortex using shunting inhibition.

The differential geometry of visual flows provides a formal expression of what "good continuation" implies for the Gestalt laws of perceptual integration. This suggests that the different flow-like visual cues, including motion, texture, and color, are all amenable to this style of analysis; but perhaps more importantly it further suggests that it is interactions between these different flows that provides the real foundation for qualitative surface inferences. While here we applied our system to shading flow fields and edges, the same ideas can be generalized to other flows—such as hue—and that the totality of their interactions should provide even richer foundations for more general scene properties.

References

- Ben-Shahar, O., & Zucker, S. (2001). On the perceptual organization of texture and shading flows: From a geometrical model to coherence computation. In *Proceedings of Computer Vision and Pattern Recognition*, 1048–1055.
- Ben-Shahar, O., & Zucker, S. (2003). Geometrical computations explain projection patterns of long-range horizontal connections in visual cortex. *Neural Computation*, 16 (3), 445–476.
- Ben-Shahar, O., & Zucker, S. (2003). The perceptual organization of texture flow: A contextual inference approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25, 401–417.
- Ben-Shahar, O., & Zucker, S. (2004). Hue geometry and horizontal connections. *Neural Networks*, 17, 753–771.
- Borg-Graham, L. Monier, C., & Fregnac, Y. (1998). Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature*, 393, 369–373.

- Bosking, W. Zhang, Y. Fitzpatrick, S. B., & D. (1997). Orientation selectivity and the arrangement of horizontal connections in the tree shrew striate cortex. *Journal of Neuroscience*, 17 (6), 2112–2127, March 15.
- Breton, P., & Zucker, S. (1996). Shadows and shading flow fields. In *Proceedings of Computer Vision and Pattern Recognition*.
- Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8 (6), 679–698.
- Dobbins, A. Zucker S., & Cynader M. (1987). Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature*, 329 (6138), 438–441, October.
- Gibson, J. (1950). *The Perception of the Visual World*. The Riverside Press.
- Gilbert, C. (1992). Horizontal integration and cortical dynamics. *Neuron*, 9, 1–13.
- Hubel, D., & Wiesel, T. (1977). Functional architecture of macaque monkey visual cortex. In *Proceedings of Royal Society B: Biological Sciences*, volume 198, 1–59.
- Huggins, P., & Zucker, S. (2001). Folds and cuts: How shading ows into edges. In *International Conference on Computer Visions, IEEE Computer Society*.
- Hummel & Zucker, S. (1983). On the foundations of the relaxation labeling proceeses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5, 267–287.
- Iverson, L., & Zucker, S. (1995). Logical/linear operators for image curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17 (10), 982–996.
- Kanizsa, G. (1979). *Organization in Vision: Essays on Gestalt Perception*. Praeger Publishers.
- Kass, M., & Witkin, A. (1987). Analyzing oriented patterns. *Computer Vision, Graphics, and Image Processing*, 37, 362–385.
- O'Neill, B. (1966). *Elementary Differential Geometry*. Academic Press.
- Parent, P., & Zucker, S. (1989). Trace inference, curvature consistency, and curve detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11 (8), 823–839.
- Perona, P. (1998). Orientation diffusion. *IEEE Transactions on Image Processing*, 7 (3), 457–467.
- Rao, A., & Jain, R. (1992). Computerized flow field analysis: Oriented texture fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17 (7), 693–709.
- Stevens, K. (1983). The line of curvature constraint and the interpretation of 3d shape from parallel surface contours. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 1057–1061.
- Tang, B. Sapiro, G., & Caselles V. (2000). Diffusion of general data on non at manifolds via harmonic maps theory: The direction diffusion case. *International Journal of Computer Vision*, 36 (2), 149–161.
- Todd, J., & Reichel, F. (1990). Visual perception of smoothly curved surfaces from double-projected contour patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 16 (3), 665–674.

PART II

Depiction of visual spaces

Pictorial space, a modern reappraisal of Adolf Hildebrand

Jan J. Koenderink and Andrea J. van Doorn

“Pictorial Space” is the space a human observer becomes aware of when looking “into” a picture as opposed to looking “at” a picture (Gombrich 1960). Suppose the picture is a conventional postcard with a photograph of the Parthenon. Then, when you look *at* the picture, you see a rectangular, flat piece of paper covered with pigments in a certain spatial order.¹ This object is part of the actual scene in front of you. When you look *into* the picture, you see a scene that is in some respects similar to what you would see in front of you when geographically displaced to some particular site near Athens. That space is not part of the scene in front of you, as is indeed clear from the fact that a similar event occurs when you look at some phantasmic scene on the cover of a science fiction novel. It has nothing to do with the rectangularity or flatness of the postcard, nor with the spatially distributed pigments on its surface, for in this *pictorial space* the postcard, as a visual object in the scene in front of you, doesn’t exist. The pictorial space is extended both laterally (somehow mirroring the two dimensions of the picture plane) and in depth, that is to say, it contains objects at various (apparent) distances from you. These “pictorial objects” are seen as extended in a three-dimensional space, as having well defined positions and spatial attitudes with respect to you and to each other, and as possessing material properties such as color, texture, metallicness, glossiness, translucency, and so forth. The light field in the pictorial scene is well defined and totally unrelated to the light field in the scene in front of you. All this should be thoroughly familiar.

Thus there exists a complete schism between the visual world (the scene in front of you) and pictorial space.² Perhaps most interestingly, you are simultaneously aware of *both* of these worlds. Usually you will treat your visual world as reflecting the structure of the physical world of which you happen to be part at this very moment, whereas you are aware of the fact that the pictorial world, despite *looking absolutely real*, is disconnected from that physical world. For instance, you are quite aware of the fact that you will never be able to *touch* pictorial objects,

although a pictorial object may well “look soft” (to the touch). The fact that the pictorial world exists only as a thread of awareness doesn’t necessarily bother you. Yet this seems to be a singularly *human* trait. With the possible exception of some of the primates, animals look *at*, not *into* pictures (van de Grind 1997), or, what amounts to the same thing, postcards are no *pictures* to them.

The issue of veridicality

The recognition of the fact that there is no *necessary* connection between pictorial objects and any physical objects is crucial to the scientific study of pictorial space. Yet the majority of studies one encounters in the literature ignore this fundamental insight. Many studies even concentrate singularly upon the “veridicality” of pictorial perception, that is on the congruity of pictorial perceptions and physical objects. This arises from a firm, but mistaken, conviction that pictures can be understood as representative records of physical objects (Marr 1982). In order to make this clear we proceed to discuss some thought experiments.

Consider the simple scenario that Alice takes a photograph of some object (e.g., the Nike of Samotrace at the Louvre) and shows it to Bob. Then Alice interrogates Bob (perhaps subjects Bob to formal psychophysical experiments) and compares Bob’s responses to her (Alice’s) description of the physical object (the actual stone Nike at the Louvre). To the degree that the responses and the parameters of the object are isomorphic, Bob’s perceptions are evaluated as more or less “veridical” by Alice. Here we assume that Bob never saw “the real Nike”.

This is probably the setting that naturally occurs to people and indeed probably the only setting in which the notion of “veridicality” can be construed to make sense as applied to Alice’s interpretation of Bob’s perceptions (not as applied to Bob’s perceptions *per se*.) There are many obvious variations, e.g., Bob may be familiar with the Nike, Bob and Alice may be the same person, the photograph may have been selected by Alice from a book (she never having seen the Nike herself), In neither of these variations can the notion of veridicality be construed to make sense though.

Consider a variation of this scenario in which the picture is a painting of the ever popular “Saint Christopher”. Since “Saint Christopher” has been officially declared by Rome never to have existed (having been removed from the calendar in 1969),³ there is indeed no way “veridicality” can be verified. What could “veridicality” mean in this case? Perhaps conformity of Bob’s perception to the painter’s mental image (whatever that might be) at the time of execution of the painting?

Yet another scenario concerns a postcard-sized piece of paper found on a garbage heap and overgrown with dark fungus so as to define a certain spatial distribution of tone. Although no “picture” (not a photograph, not a painting, ...),

Bob “sees” a phantasmic “pictorial space”, say a dance of elves in the moonlight. Clearly no notion of “veridicality” applies here. Yet such cases are indeed common enough (you often see “faces” in clouds, and so forth.).

What is the difference between a photograph of the Nike of Samotrace and a piece of paper overgrown with fungus that (*mirabile dictu*) has exactly the same spatial distribution of gray values? Bob’s pictorial spaces for the two objects will of course be identical. The physical difference (fungus or silver grains in gelatin) is irrelevant if you believe (as most people would) that a printed postcard of the Nike and a photograph of the Nike are equivalent as *pictures* though different as physical objects (printing ink spots versus silver grains in gelatin). If you grant this then the “veridicality” as verified by Alice in the first example exists only relative to *her own* awareness and experience and is quite irrelevant as applied to *another person’s* (that is Bob’s) awareness.

This is typical for most of the literature: Although the “veridicality” does evidently not apply to the *subject* in the experiment (the observer), but to the *authors* (the person who prepared the picture and ran the experiment), this goes apparently unrecognized which again leads to various exercises in nonsense.

The psychophysics of pictorial perception

With “veridicality” discarded as a non-issue, what kinds of questions *can* be addressed in “pictorial perception” in an operationally and conceptually *well* defined manner?

There is no doubt that pictorial perceptions depend at least partly on the structure of the *picture*. Few people will perceive an elves’ dance in a straight photograph of the Nike of Samotrace. This opens up a huge field of enquiry. You can vary a picture parametrically and look for covariations in the perception (for example Koenderink & van Doorn 2001).

There is also no doubt that pictorial perceptions will depend upon the *perceiver*. For instance, it seems established that cats see nothing in a picture of the Nike although these animals are keen visual perceivers. The late Sir Ernst Gombrich (1960) referred to this as “*The Observer’s Share*”. This opens up a huge field of enquiry too. You can compare Bob’s perception to Alice’s on the same picture (for example Koenderink & van Doorn 2001).

There is also no doubt that the *viewing conditions* (monocular, binocular, oblique versus frontal, ...) codetermine pictorial space. This also opens up a huge field of enquiry, easily amenable to standard psychophysical methods (for example Koenderink & van Doorn 2001).

Thus ignoring the issue of veridicality by no means obviates a *science of pictorial space*.

How would you *measure* properties of pictorial objects, or pictorial space? Think of properties such as mutual distances between objects, relative size of objects, spatial attitudes of surface elements, aspects of shape such as curvatures, material properties such as translucency, glossiness or spectral signature (“proper color”), direction and diffuseness of local illumination, and so forth. Here numerous potential methodologies come to mind. By way of a specific example we suggest just one generic paradigm that we have found of recurring use ourselves.

Notice that many physical measurements can be put in the form of a comparison of some substructure of the world with some conventional “gauge object”. Usually one tries to keep gauge objects particularly simple and tries to reduce the comparison to the judgment of coincidence of a feature of the substructure with a fiducial mark on the gauge figure (Eddington 1928), or even the coincidence (“fit”) of the substructure with the gauge figure as a whole. An example of the former would be a graduated yardstick, a ticking clock, and so forth. Examples of the latter kind include the use of a straightedge to check right corners, a rule to check the planarity of a table top, compasses to check the sphericity of billiard balls, a template to check screw threads, and so forth.

The virtue of this generic method is that it can be performed fully in pictorial space, for you can put a *picture* of the gauge object in pictorial space and use it as a *pictorial gauge object*. You need only compare pictorial objects among themselves, consequently no conflict of different strata of reference (e.g., objects of the physical world versus objects of the pictorial realm) need occur. This generic method can be implemented for many different purposes and in many different ways (for example Koenderink & van Doorn 2001).

The nature of pictorial cues

“Pictorial cues” are conventionally distinguished sharply from “depth cues” (Palmer 1999). “Depth” is the radial distance from the egocenter, it supposedly runs from zero (though the physical location of the egocenter is ill defined) to infinity. “Depth” is “veridical” if it is (approximately) equal to (or “proportional to”, or “monotonic with”, it all depends on your convictions) the radial *distance* to the midpoint of your interocular segment. The distance is a *physical* parameter whereas the depth is a *perceptual* entity. In much of the literature “depth” is almost synonymous with “binocular vergence”, or “binocular disparity” or (weakly) eye lens accommodation (Howard & Rogers 1995). Bishop Berkeley’s skepticism (Berkeley 1709) regarding such notions applies fully today. His ideas on cues are also remarkably useful.

Berkeley regards cues as essentially *arbitrary* associations, which – from a biological and evolutionary perspective – should be a matter of course by modern

standards (Riedl 1982; Lorenz 1965). As Berkeley understood quite well, many cues can be understood in terms of causal relations of physics and generic structure of the observer's biotope. The former has led to a large literature in "computer vision", the latter to a large literature (e.g., see Palmer 1999) in "The Bayesian Theory of Perception" (the good Reverend Bayes is entirely innocent of this). In either case these "pictorial" (as opposed to "physiological") cues depend crucially on the nature of prior assumptions, i.e., "The Observer's Share" (Gombrich 1960). Because of that the pictorial cues are considered second rate *ersatz* crutches (referred to as "secondary cues") as compared to the real thing (triangulation via vergence angle for instance: referred to as "primary cues").

Pictorial space intrinsically depends on these supposedly secondary cues, there being no alternative. (For instance, the vergence angle merely tells you that you're looking at a *flat* postcard.) Yet pictorial space is hardly less *convincing* than "real" visual space. Indeed, as was discovered in the nineteenth century (when someone by accident put two *identical* photographs in a stereoscope instead of a left-right pair) the perceptual space due to the secondary cues appears – if anything – *more real* than the space primarily due to the primary cues. Because this is so well known to be *impossible* the (perhaps unfortunately often confirmed) phenomenon became known as "paradoxical monocular stereopsis", and was only published in obscure periodicals, the major scientific journals knowing better than to compromise themselves. The topic has vanished from the modern textbooks. We recognize it as a remarkable and important empirical fact though.

When you look at a real stereo pair through a Victorean stereoscope (a city scene say) you experience a "coulisses space", that is to say, you will see great depth, objects being separated by obvious depth gaps, but the objects themselves appear (perhaps disappointingly so) totally *flat* instead of "rounded" as objects in the visual world (the scene in front of you) typically do. They look like card board cut outs. Such observations make it seem very likely indeed that the visual "realness" of objects in the visual world derives largely from the secondary (pictorial) cues, even though this goes against the grain of official wisdom.

Phenomenologically, the nature of pictorial "depth" is different in kind from a depth that is like a distance from the egocenter. In many respects the egocenter indeed belongs to the visual world (a singular location just outside the boundary of the scene in front of you) but not to pictorial space. Pictorial depth has no natural origin. Unlike the visual world there seems to be no "potential" part of the pictorial world that lies behind your back. Pictorial objects also have rather ill defined *sizes*. In some cases a pictorial scene may strike you as smaller than usual (like a doll's house), but typically you are not bothered by such feelings. Absolute size hardly applies to pictorial space. Introspectively you are simultaneously both "inside" and "outside" of pictorial space.

Various features in pictorial space can become integrated in pictorial surfaces, or pictorial relief and acquire more or less definite depth relations. The depth relations of features that live in different pictorial reliefs remain largely ambiguous. This ambiguity can only be lifted through relations to more global surfaces (also pictorial reliefs) such as the apparent groundplane. It has been an understanding of these basic facts that made apparently “realistic” landscape paintings possible. In early attempts at landscape painting the foreground, middle ground and background regions typically remain unrelated. Often the awkward transitions are masked through “artistic fig leaves” such as occlusions by conveniently located bushes or minor hill crests. Fore, middle and background then appear as coulisses at mutually undeterminate separations in depth.

Adolf Hildebrand’s “relief space”

The German sculptor Adolf Hildebrand became interested in the theory of perceptual spaces due to his interactions with the painter Hans von Marees and the philosopher and theoretician of art and aesthetics Konrad Fiedler (1960). His influential treatise “On the Problem of Form” (G: *Das Problem der Form*) started out as a discussion of Marees’ understanding of art (after the latter’s death) and was developed into a full fledged book under Fiedler’s intellectual pressure (Hildebrand 1893).

The Problem of Form is remarkable in many ways, not in the least because Hildebrand (a sculptor) treats aesthetic perception as largely pictorial perception. This was indeed considered as paradoxical by most contemporary art theoreticians. Nevertheless the book became very influential in the late nineteenth and early twentieth century art theory. Its influence can be traced in developments of art nouveau painting and eventually cubism.

Hildebrand understands visual space in so far as it is truly *optical* (he differentiates between a motor space (“Bewegungsvorstellung”) and optical space (“Fernbild”)) as *relief space*. This relief space has all the traits of pictorial space. Hildebrand conceives of “depth” as of a “flow” away from the observer and (in pictures) behind the picture surface. This flow has no well determined origin and potentially flows towards infinity (the horizon, at no particular depth). As this flow is checked, or arrested, by pictorial surfaces it reveals “pictorial relief”. Pictorial relief has shape (curvature, spatial attitude) but no well defined depth extent and various reliefs are staggered in depth like coulisses (again, at no particular relative depths). The depth *order* is well defined, but the depth dimension in a quantitative sense remains highly volatile. A key example is relief sculpture compared with sculpture “in the round”. As Hildebrand noticed the observer often remains unaware of the difference. Pictures are simply the limiting case of completely flat-

tened relief. Pictorial relief is entirely *optical*, free of any motor component such as binocular vergence or lens accommodation. The pictorial depth dimension has no physical counterpart but is entirely a thing of the mind. The beholder expands or shrinks the depth dimension idiosyncratically. It is the task of the artist to provide the optical (pictorial) cues and “composition” is a matter of design in the various “depth layers” and the artful interrelation of these otherwise disparate layers through devices *in the picture plane*. The artist attempts to guide the beholder’s share as it were, and the beholder’s task is to read these pictorial cues sympathetically, artfully exploiting the freedom left by the inherent ambiguity of the artist’s construct.

The interplay between the ambiguity left by the cues and the beholder’s share is an interesting issue and the remainder of this contribution will largely focus on this topic. Notice that – in case vision “respects the evidence of the cues” fully (which should not be confused with “veridicality”) – *the room left for the beholder’s share is exactly what is left undetermined by the cues*. Thus the beholder’s share cannot be *fully* idiosyncratic, that is to say, the only irreducibly idiosyncratic part is due to the beholder’s “understanding” (not in the cognitive sense of course) of the cues.

The degree to which the perceiver may use the structure of the pigmentation of the picture surface *defines* what are cues for that observer. A world without visual observers would contain no cues. If Berkeley is right that cues are learned through arbitrary association (as he surely is), then we expect the expertise to use the cues to be similar for observers raised in similar biotopes (Lorenz 1965; Riedl 1982).

We conclude that the core of “the beholder’s share” is *carved out by the cues* so to speak. Since the cues can be understood from the perspective of ecological physics (despite their being learned by arbitrary association), an understanding of the beholder’s share should be possible from an analysis of the optics (physics of the biotope). (See books on “Computer Vision”, e.g., Forsyth & Ponce 2002.) As we will argue later, this offers a handle on the formal understanding of the structure of pictorial space.

Hildebrand defines the beholder’s share largely as depth contractions and dilations. He (at least implicitly) assumes that such transformations conserve projective structure (lines, planes, and their structure of intersection). This allows one to write down the formal structure of the group of transformations that leave the structure of configurations in pictorial space intact, you may understand it as the group of congruences, or proper movements. Following Felix Klein’s “Erlangen Program” (1872) this again *defines the geometry* of pictorial space.

Since Hildebrand never performed formal psychophysical investigations (though he produced a number of demonstrations that may serve as “instant psychophysics”), a thorough investigation of the extent of the beholder’s share, or the group of congruences of pictorial space, is in order. Hildebrand also refrained from theoretical considerations of (ecological) physics. Thus a thorough investigation of

the group of ambiguities left by the cues is in order too. One would expect these theoretical investigations to “explain” (or, perhaps more modestly, “describe”) the results of the psychophysical investigations. Together they would yield a formal and empirically underpinned understanding of the structure of pictorial space.

Although such a program should have the highest priority in the study of pictorial vision it has never been even attempted. We will describe a number of empirical results and some preliminary theoretical considerations that allow a first glimpse of what a structured understanding of pictorial vision might look like.

Some empirical observations

Using psychophysical methods discussed above, we can use a *picture* as psychophysical *stimulus* and measure an (operationally defined) *pictorial relief* as

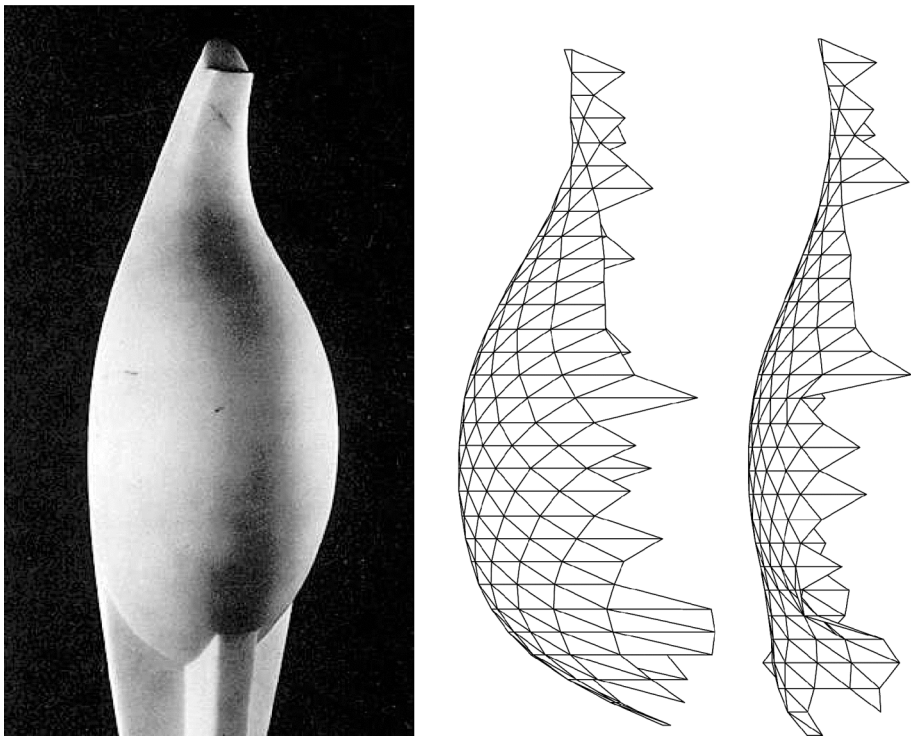


Figure 1. At left the stimulus. It is a photograph of Brancusi's "Bird" at the Philadelphia art museum, which none of the observers had actually seen. At right the responses of two observers, rotated such that the depth dimension runs from left to right. Notice that these are not congruent, thus at least one of the responses "fails to be veridical".

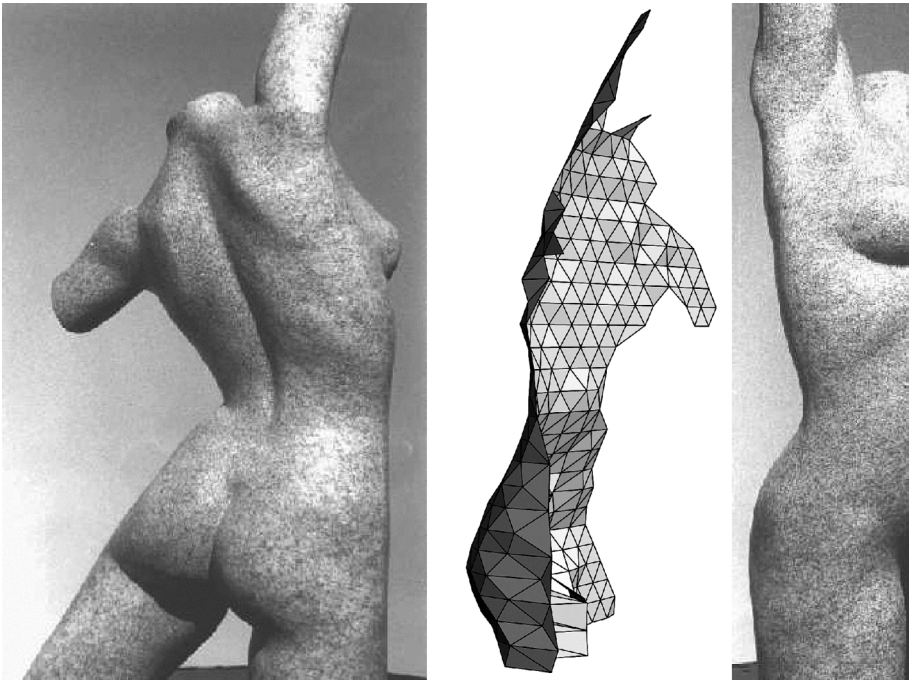


Figure 2. At left the stimulus. This picture is the straight photograph of a plastic torso seen in dorsal view. At center the response, this is the profile view of the pictorial relief, the depth dimension running from left to right. At right another photograph of the same torso, but rotated (in physical space of course) over 90° with respect to the photograph at left. Thus the contour in this photograph can immediately be compared to the contour in the response figure. Notice that these curves are clearly not congruent, although there is a certain family likeness between them. This does not necessarily indicate that the response “fails to be veridical”, for there exist infinitely many (virtual scenes) that would produce exactly the same photograph. Thus the stimulus does not uniquely imply a certain “veridical response”.

“response” (Koenderink & van Doorn 2001). We will show a few examples where the observer’s task was to sample local (pictorial) surface attitude. From a sufficiently dense set of such samples we can construct an approximate pictorial relief. The pictorial relief is a surface in three dimensional space, one of whose dimensions is “depth”. By plotting a “profile” of the pictorial relief we look immediately at the *depth* “seen” (again, in this operational sense) by the observer.

In view of the literature it is perhaps opportune to demonstrate that pictorial relief is in general not “veridical” in the naïve sense. In Figure 1 we show profiles of pictorial reliefs of two observers for the same picture. Being obviously different, at least one is not veridical in the naïve sense. In Figure 2 we show the profile of

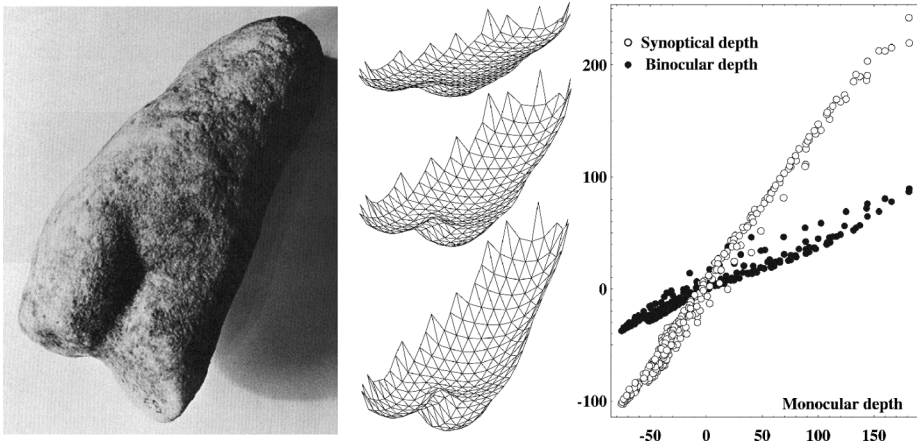


Figure 3. At left the stimulus. It is a photograph of an eroded piece of Greek sculpture. At the center three responses of the same observer, oriented such that the depth dimension runs from bottom to top. The upper response was obtained with binocular vision, the middle one with monocular vision (the observer closed one eye), the bottom one with synoptical vision. In synoptical vision both eyes are open, but the two retinal images are artificially caused to be identical. The responses have been drawn on the same scale. At right the scatter plot of depth values at the vertices of the triangulation shown in the responses. The relation of binocular depth to monocular depth is linear and so is that between synoptical depth and monocular depth. Thus we have a pure case of depth dilation or contraction as described by Adolf von Hildebrand.

pictorial relief for one observer with the actual (physical) profile. Since they are different, the relief is not veridical in the naïve sense.

Pictorial relief depends upon viewing conditions as artists have known for centuries. It is possible to quantify this. In Figure 3 we demonstrate that for some particular observer the total depth range for monocular viewing is about twice the total depth range for binocular viewing, neatly illustrating Leonardo da Vinci's observations (MacCurdy 1939). The depth range again expand by roughly a factor of two for "synoptical viewing" (identical input to the eyes, as when viewing two identical pictures in a stereoscope).

One notable observation from the above data is that Hildebrand's ideas concerning dilations and contractions of the depth domain are indeed very common and quantitatively remarkably large. Factors up to (and exceeding) five are commonly encountered.

Even more remarkable are the observations shown in Figure 4 (Koenderink et al. 2001). Here a single picture was sampled by the same observer in somewhat different settings (it is not particularly important here what these differences were). The reliefs cannot be put into depth correspondence via a Hildebrand depth con-

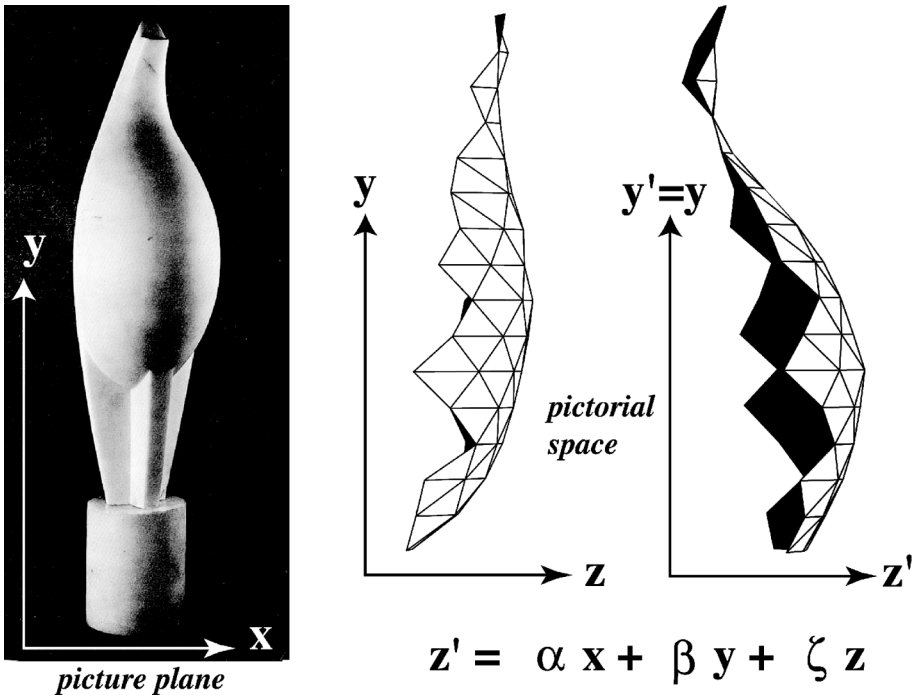


Figure 4. At left the stimulus. It is a photograph of Brancusi's "Bird" at the Philadelphia art museum, which the observer had never actually seen. At right two responses, obtained for the same stimulus (the photograph at left), though under slightly different circumstances. The responses are oriented such that the depth dimension runs from left to right. Notice that the responses look similar, though different. It is as if one had been "rotated in pictorial space" with respect to the other. Quantitative analysis reveals that it is not a Euclidean rotation though, but that the relation is a non-Euclidian movement described by the formula below. It is a congruence of isotropic space. The movement that relates the responses is clearly not specified by any pictorial cue. It has to be fully idiosyncratic and might well be referred to as "a mental movement (or change of perspective) in pictorial space".

traction or dilation. We find that excellent correspondence is obtained when not a depth scaling, but a certain *shear* in pictorial space is applied. The shears that work leave the locations in the picture plane invariant, notice that these locations are (mutually parallel) *lines* (extended in depth) in pictorial space. This class of transformations was not (at least not explicitly) identified by Hildebrand. However, we have encountered it again and again in our experiments, it has become increasingly evident that they are on the same footing as the original Hildebrand dilations and contractions: They are to be reckoned among the *proper movements* or *congruences* of pictorial space.

A preliminary theory of the optical basis of cue ambiguity

Many of the conventional “pictorial” (or “monocular”) cues have been subjected to extensive formal analysis in “Computer Vision” (Forsyth & Ponce 2002). It has become common understanding that any of these cues is problematic in the sense of depending on uncertain prior assumptions, and that – even granted these assumptions are satisfied – the “solutions” are not unique, but represent infinite sets of equally admissible solutions. Typically there exist groups of transformations (Koenderink & van Doorn 1997; Belhumeur et al. 1999) that transform one possible solution into other possible solutions. These transformations typically have a group structure and depend upon a number of free parameters. We will refer to such groups as “ambiguity” groups.

Sometimes the ambiguities of different cues are different, in that case the cues together define a smaller group of ambiguities, the intersection of the original groups. It has become more and more evident, that there is little hope to expect the final intersection (putting into effect more and more different cues) will verge to the identity (no ambiguity) though. In fact, many of the better understood cues have *very similar* (largely overlapping) ambiguity groups. Thus the future looks grim for a robot whose (veridical) vision has to rely on pictorial cues. Engineers look for alternatives, such as active probes (ultrasound echo locators and so forth.) to ameliorate this situation.

The group of proper movements of pictorial space

There is little hope to have a complete, formal description of all (humanly) relevant pictorial cues in place in anything like the near future. By way of a cheap substitute we present a very simple argument that will at least get us in the right ball park. Notice that all of the cues can only move points back and forth in depth, for the locations in the visual *field* are quite certain and independent of any cues. This means that the ambiguities should conserve a family of mutually parallel lines as a family. We *assume* that

- pictorial space is a *homogeneous* space, that is to say, a space that looks the same from any of its points. This can be regarded as an assumption of least commitment or explanatory efficiency.

We will also *assume* that

- the relevant cues are at best able to signal deviations from planarity.

This is an uncertain assumption, though it is rendered somewhat reasonable by the fact that at least the better known cues are all of this variety. It is the only assumption of real *substance*.

Remarkably, once these two assumptions are accepted we are done. For the only homogeneous space that satisfies the assumptions is one of the twenty-seven Cayley–Klein spaces. From a technical point of view it is the one with one “isotropic dimension” (depth), the planes containing this dimension having parabolic distance and angle metrics, the planes orthogonal to this direction being Euclidean (i.e., essentially the picture plane, with a parabolic distance but an elliptic angle metric).

The group of similarities of this space is an eight-parameter group,⁴ one degree of freedom more than the group of similarities of the Euclidean space that describes the physical world. This derives from the different angle metric. In Euclidean space distances can be scaled but angles cannot, because they are periodic. In planes containing the isotropic direction angles are not periodic and can be scaled, leading to a type of similarity unknown to physical space. This is the reason for the additional parameter.

The geometry of pictorial space

The geometry of pictorial space is perhaps most easily understood by first considering the action of motions and similarities confined to certain special planes.

First consider the planes parallel to the picture plane. In these planes the geometry is simply Euclidean. The proper motions are translations and rotations about arbitrary angles with the understanding that angles are periodic. A translation adds some arbitrary vector to all location vectors and thus doesn’t change distances, nor angles or directions. A rotation adds some arbitrary angle to all directions, but does not affect distances, nor angles. There is one type of similarity, a uniform scaling, changing all distances by the same factor and leaving angles invariant. In Euclidean planes the distance metric is parabolic (e.g., the x -coordinate ranges between minus and plus infinity), whereas the angle metric is elliptic (e.g., directions range from 0 to 2π).

The geometry in planes that contain the depth direction is very different from Euclidean (Yaglom 1979; Strubecker 1938; Sachs 1987 and 1990). (Notice that such *planes* of pictorial space appear as *lines* in the picture plane because the depth dimension is not represented in the picture plane.) The proper motions include translations, exactly like in the Euclidean plane. One component of such a translation is a *depth shift*, which may be considered irrelevant because absolute depth doesn’t really mean anything in pictorial space. (The depth dimension having no true origin.) The other component is a *rotation*, because of the fact that the plane

contains an isotropic direction (depth) often called “isotropic rotation”. Isotropic rotations look like shears to the “Euclidean eye”. Such shears conserve an isotropic line, which is the “center of rotation”. All other isotropic lines suffer depth shifts of various amounts, proportional to their distance in the picture plane from the center of rotation. Such a depth shift, divided by the distance, is the *isotropic angle* of rotation (it is indeed the same for all points). The isotropic angle of rotation varies between minus and plus infinity, that is to say, rotations are *not periodic*.

Thus objects in pictorial space *cannot make a full turn*. This is indeed a necessity, because pictorial objects only exist due to an invariant distribution of pigments on the picture surface. A frontal portrait will never show the back of the head, even though the beholder may (isotropically) rotate the frontoparallel plane. No such “turn” will ever reveal features that are not “in the picture”. This explains part of the impact of Magritte’s painting “The Schoolmaster” (painted 1954) which shows a “portrait” painted from “the wrong side” (showing only the back of the head). The schoolmaster is literally faceless and no amount of mental rotation will ever yield the (perhaps unconsciously) desired *en face* view. This simple fact alone should suffice to demonstrate *ad oculos* that pictorial space cannot be Euclidean.

The proper motions leave the distance in the picture plane invariant, thus this distance is a prime candidate for “the distance” of the isotropic plane. Indeed, this often works quite well. For instance, one sees that a “unit circle” consists of two isotropic lines, with the isotropic line at equal distance from each as center. Rotations indeed move such (isotropic) circles along themselves, the center remaining fixed. There appears to be a problem though: Two points may be at zero mutual distance, but nevertheless be *distinct*. We cheerfully accept this oddity and call such points mutually *parallel*. If lines can be “parallel”, why not points? For parallel points the depth difference can be used as a substitute distance measure, for it is easily verified to be invariant under proper motions. Thus one defines “the distance” as *either* the distance in the picture plane, *or* (if the points are parallel) the depth difference. With this definition the (isotropic) rotation angle is simply the arc length along the (isotropic) unit circle. We have obtained an internally consistent, non-Euclidean geometry that has both remarkable similarities and remarkable differences from the trusty Euclidean plane.

Isotropic geometry is in many respects similar to, and every bit as “rich” (in terms of aesthetically pleasing theorems) as the geometry of the Euclidean plane (Yaglom 1979; Strubecker 1938; Sachs 1987, 1990). In some respects it is even aesthetically *more* pleasing, because more “symmetric” than the Euclidean plane though. In both Euclidean and isotropic planes there exists a remarkable symmetry between points and lines (in space between points and planes). This has led to the development of so called “projective geometry”, which is a subset of Euclidean geometry in which the symmetry is *perfect* (Hilbert & Cohn-Vossen 1999). In the

Euclidean plane the symmetry remains flawed though. For instance, whereas it is indeed generically the case that “two distinct points define a line” and “two distinct lines define a point”, this remarkable symmetry is not without exceptions: Whereas two parallel lines don’t define any point, two distinct points always define a line. Euclidean geometry is rife with such exceptions, whereas projective geometry has none by design (the very concept of “parallelity” is meaningless in projective geometry). Isotropic geometry also knows none such irritating exceptions, but the concept of parallelity is well defined. In isotropic geometry there exist parallel points, just as there exist parallel lines.

In contradistinction with projective geometry, the symmetry is *metrical*. Projective geometry (Birkhoff & MacLane 1996) has no two-point (or two-line) invariant, but in the isotropic plane distance and angle *are two point* invariants between which perfect symmetry rules. Thus isotropic geometry is in many respects remarkable, and indeed, Strubecker (1938) and also Yaglom (1979) recommended its teaching to youngsters, as far more rewarding than the conventional teaching of Euclid’s Elements (Byrne 1847). They were right, though perhaps less than realistic.

When we consider the geometry of pictorial *space*, rather than simply the geometry in the special planes considered above, things are rather more complicated (Strubecker 1938; Sachs 1990). Its study is rewarding though and certainly important. For instance, the intrinsic definition of “pictorial shape” has to depart from the differential geometric invariants defined by the group of proper movements and similarities of pictorial space. For “shape” is what remains the same under changes of location, attitude and size. This is an important topic, with many real world applications.

Adolf Hildebrand intuited the larger part of the structure of pictorial space. The one thing he apparently missed was the subgroup of isotropic rotations. As we have indicated earlier, these isotropic rotations appear routinely in psychophysical studies and we must consider their existence to be a well established fact. We find it very hard to imagine that Hildebrand could have been unaware of this. Indeed, we find many reasons that make it likely that he had to be. However, perhaps sadly, we failed to find any evidence of this in his written heritage.

The isotropic rotations are of much interest because they can be understood as “mental changes of viewing direction”, for – being part of the ambiguity group – they are inherently idiosyncratic, and they do change the apparent “frontoparallel” plane. Notice that the notion of “frontoparallel” is dual to the notion of “origin” in the depth dimension. This is yet another aspect of the metrical duality between distance and angle in isotropic geometry. Just as there exists no clearcut “origin”, there exists no clearcut “frontoparallel” in pictorial space. Therefore, when depth shifts are irrelevant, then changes of the apparent frontoparallel plane *should* be irrelevant too. It is our limited experience with the structure of pictorial space that

makes us find the one trivial, the other remarkable. Still, it remains the case that observers commit both frequent idiosyncratic depth shifts (“mental movements in depth”) and frontoparallel plane changes (“mental rotations”).

Conclusion

Modern psychophysics thus far has amply corroborated Adolf Hildebrand’s intuitions as expounded in his monumental *Das Problem der Form*. This work remains without doubt the most insightful treatment of pictorial space in history, and it has been rightly influential in its time, though (sadly enough) largely forgotten by now. Modern empirical data have shown that the beholder’s share is even somewhat more expansive than envisaged by Hildebrand, including (isotropic) rotations of the frontoparallel plane in addition to depth scalings. A formal analysis reveals full metric duality between points and planes in pictorial space, which puts the depths shifts and rotations in the same category of “mental movements”.

Hildebrand’s depth dilations and contractions are formally understood as *similarities* that scale the isotropic angles (“similarities of the second kind” in Strubecker’s parlance). They are dual to scalings of the distances, that is the dimensions of the picture plane. When nobody doubts that the same pictorial space will result from a postcard and a jumbo-sized copy of that same postcard, then (if one buys the formal geometry) nobody should be surprised that the dual property should also hold (the Hildebrand depth dilations and contractions). Yet many are surprised that (as Hildebrand noted) observers may be at a loss to distinguish between bas relief sculpture and sculpture in the round.

The discussion above should make it clear that the structure of pictorial space is essentially different from the structure of visual space. The egocenter is part of visual space, but not of pictorial space.⁵ Visual directions *diverge* from the egocenter, but are mutually parallel in pictorial space. The frontoparallel planes are well defined in visual space, but very ambiguous (like absolute depth) in pictorial space. And so forth. Such observations are indeed generically true, but there appear to occur occasional exceptions. Whereas the visual rays in pictorial space are almost without exception experienced as “parallel” (Note 4), the visual rays in visual space are generally experienced as diverging but to an often quite unrealistic degree. Frequently the divergence is severely underestimated, even to the extent that visual space approaches a form of pictorial space. The degree to which this occurs appears to be a personal trait, differing from person to person. In Hildebrand’s discussion of the “Fernbild” he may (also) have referred to this fact.

Since the picture – because a physical object – also appears in the visual world, it frequently happens that visual space and pictorial space contain a connecting “wormhole”, connecting two otherwise disparate worlds. In the case of a framed

picture hanging on a wall the picture frame sometimes assumes a double identity as it were, as an object of the visual field *and* as an object in pictorial space. In such a case the observer may become aware of the different nature of these spaces. For instance, there are many cases of portraits where the pictorial face “follows” with its gaze as the beholder assumes different locations with respect to the painting. (The well known world war one poster of Lord Kitchener pointing at the beholder with the text “I want YOU for the army” is a particularly compelling case. See Gombrich 1960.) Such paintings are sometimes described as “bewitched”, and thought to be due to almost supernatural abilities of the painter. Such is far from being the case. Formal psychophysical experiments (Koenderink et al. 2004) show that the visual rays – although oblique with respect to the physical picture surface and oblique to the picture surface (as defined by the frame in visual space) – remain (as isotropic rays in pictorial space) normal to the picture plane (as defined by the frame in pictorial space). The observer appears to apply the geometry of the visual world to both worlds, with the result that the congruences of the pictorial world appear as deformations in the visual world. Such observations have led to a large (and largely spurious) corpus of literature.

Notes

1. This refers to the famous (and at his time revolutionary) dictum of Maurice Denis (1890): “It should be remembered that a picture – before being a war horse, a nude, or an anecdote of some sort – is essentially a flat surface covered with colors assembled in a certain order.”
2. Exceptions occur when the picture is no longer perceived *as* a picture. This may happen when the scene in front of you is very indistinct relative to the picture, for instance in a movie theatre, or when the picture assumes such a large visual extent that it occludes much of the scene in front of you, as in the nineteenth century panoramas. A combination of these, say a very large screen movie theatre works even better still. In such cases the picture tends to assume the role of the scene in front of you. Hildebrand (1893) feels that such cases are in the worst of taste, a bit like when you would pass the wax puppets of Madame Tussaud’s for artistic sculpture. In a sense he is literally right in this.
3. According to Jacobus de Voragine’s (1275) *Aurea Legenda* Christopher was a Canaanite ogre living during the first half of the 3rd c. The *Calendarium Romanum* was revised by the instruction of the Vatican Council II and published by Pope Paul VI (2/14/69). In the list of “variations” under 25 July (the feast of St. James the Apostle) is the entry: “*the memorial of Christopher, which entered the Roman Calendar in 1550, is not a part of the ancient Roman tradition. It is now left to particular calendars. Although the Acts of the life of Christopher are legendary, the existence of his cult is very old*” (p. 86). Perhaps surprisingly, there is no declaration of non-personhood, despite the fact that de Voragine records the height of St. Christopher as twelve cubits (ca. 5.5 m).

4. The group of similarities of pictorial space is (in conventional matrix notation)

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \mu \cos \varphi & \mu \sin \varphi & 0 & \tau_x \\ -\mu \sin \varphi & \mu \cos \varphi & 0 & \tau_y \\ \sigma_x & \sigma_y & \gamma & \tau_z \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

where the $\{x, y\}$ -coordinates span the picture plane and the z -coordinate designates the depth dimension. Notice that the picture plane is Euclidean, with rotations over angles φ (which is periodic), translations by vectors (σ_x, σ_y) , and a uniform scaling by factors μ . The depth transforms as

$$z' = \sigma_x x + \sigma_y y + \gamma z + \tau_z$$

with (to the Euclidean eye) “shears” in the xz and yz -planes, which are actually isotropic “rotations” over angles $\sigma_{x,y}$ (not periodic), a scaling with factor γ , and a depth shift τ_z . The isotropic scalings and depth shifts are the transformations described by Hildebrand, whereas the isotropic rotations are routinely encountered in psychophysical experiments.

Notice that the $\{x, y, z\}$ coordinates are best understood as the $\{u, v, w\}$ coordinates introduced in Note 5. Especially, the z -coordinate should be understood as an affine line, ranging between minus and plus infinity, and lacking a natural origin.

5. Indeed, the apparent angular extent of the pictorial space is indeterminate. Introspectively the visual directions seem to be parallel to each other and the position of the egocenter is somehow outside pictorial space. A simple model that accounts for this is the logpolar model. In the two dimensional case (one dimensional picture) points can be denoted by polar coordinates $\{r, a\}$, where r denotes the depth, and a the visual angle. The logpolar transformation $\{r, a\} \rightarrow \{u, w\}$, with $u = \log r/r_0$ and $w = a$ (for some reference distance r_0) is a conformal transformation that puts the egocenter ($r = 0$) at minus infinity, i.e., outside the space and transforms all visual rays to a pencil of parallel lines. This has several advantages. It explains why the depth dimension has no natural origin (r_0 being arbitrary), why the visual rays don't diverge, and why the egocenter cannot be in the space. For not too large values of the maximum value of a (the common case) this can be generalized to the three dimensional case (you now get two coordinates $\{u, v\}$ instead of the single u , w remains the same). The picture plane can be understood as the $\{u, v\}$ plane, in arbitrary units. This explains why people typically refer to wide angle and tele lenses as “distorting space”. This “distortion” is apparently with respect to an absolute notion of “normal” visual space. The observers are unable to allow for the change in divergence of the photographic rays at the time of exposure and that of the visual rays at the time of viewing the picture. This mismatch produces the feeling of distortion, the lenses being (in general) distortion free.

References

- Belhumeur, P. N., Kriegman, D., & Yuille, A. (1999). The Bas-Relief Ambiguity. *International Journal of Computer Vision*, 35(1), 33–44.
- Berkeley, G. (1709). *New Theory of Vision*. Dublin.
- Birkhoff, G., & Mac Lane, S. (1996). Projective Geometry. In *A survey of modern algebra*, 5th. ed. (pp. 275–279). New York: Macmillan.

- Bock, H. (1988). Adolf von Hildebrand: Gesammelte Schriften zur Kunst, bearbeitet und mit Einführung versehen von Henning Bock. *Wissenschaftliche Abhandlungen der Arbeitsgemeinschaft für Forschung des Landes Nordrhein Westfalen*, Bd. 39, Köln und Opladen 1969, 2. Auflage.
- Byrne, O. (1847). *The Elements of Euclid*. London: William Pickering.
- Denis, M. (1890). *Manifesto of Symbolism*.
- Eddington, A. S. (1928). *The Nature of the Physical World*. MacMillan.
- Fiedler, C. (1960). *Über die Beurteilung von Werken der bildende Kunst*. Ratingen (Rhld.): Henn.
- Forsyth, D., & Ponce, J. (2002). *Computer Vision: A Modern Approach*. Upper Saddle River, NJ: Prentice Hall.
- Gombrich, E. (1960). *Art and Illusion*. London: Phaidon.
- Grind, W. A. van de (1997). *Natuurlijke intelligentie*. Amsterdam: Nieuwezijds.
- Hilbert, D., & Cohn-Vossen, S. (1999). *Geometry and the Imagination*. New York: Chelsea.
- Hildebrand, A. (1969). *Das Problem der Form*. 1893, see Bock, Adolf von Hildebrand: Gesammelte Schriften zur Kunst.
- Howard, I. P., & Rogers, B. J. (1995). *Binocular Vision and Stereopsis*. Oxford: Oxford University Press.
- Klein, C. F. (1872). *Vergleichende Betrachtungen über neuere geometrische Forschungen* (the “Erlanger Programm”). Erlangen, Andreas Deichert.
- Koenderink, J. J., & van Doorn, A. J. (1997). The generic bilinear calibration-estimation problem. *International Journal of Computer Vision*, 23(3), 217–234.
- Koenderink, J. J., & van Doorn, A. J. (2001). Pictorial space. In H. Hecht, R. Schwartz, & M. Atherton (Eds.), *Looking into pictures. An Interdisciplinary Approach to Pictorial Space* (pp. 239–299). Cambridge, MA: The M.I.T. Press.
- Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., & Todd, J. T. (2001). Ambiguity and the “mental eye” in pictorial relief. *Perception*, 30, 431–448.
- Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., & Todd, J. T. (2004). Pointing out of the picture. *Perception*, 33, 513–530.
- Lorenz, K. (1965). *Evolution and Modification of Behavior*. Chicago: University of Chicago Press.
- MacCurdy, E. (1939). *The Notebooks of Leonardo da Vinci*. New York: Reynal & Hitchcock.
- Marr, D. (1982). *Vision*. San Francisco: Freeman.
- Palmer, S. E. (1999). *Vision Science*. Cambridge, MA: The M.I.T. Press.
- Riedl, Rupert (1982). *Evolution und Erkenntnis*. München.
- Sachs, H. (1987). *Ebene Isotrope Geometrie*. Braunschweig/Wiesbaden: Friedr. Vieweg & Sohn.
- Sachs, H. (1990). *Isotrope Geometrie des Raumes*. Braunschweig/Wiesbaden: Friedr. Vieweg & Sohn.
- Strubecker, K. (1938). Beiträge zur Geometrie des isotropen Raumes *J.f. die reine und angewandte Mathematik*, 178, 135–173.
- Voragine, J. de. *Area Legenda*, compiled by Jacobus de Voragine, 1275, Englisied by William Claxton, 1483. Numerous modern editions.
- Yaglom, I. M. (1979). *A Simple Non-Euclidean Geometry and its Physical Basis*. New York: Springer.

Gestalts of thought*

Barbara Tversky

Perceptual organizing principles: Grouping

Gestalt psychology is by now a given, recognized by the public at large, accepted by the scientific community, immortalized in dictionaries as “the whole is more than the sum of the parts.” The Gestalt approach developed to describe certain compelling phenomena of the experience of perception that are not easily explained by qualities of the stimuli nor of the internal processing of them. Most prominent among them is grouping. The mind groups visual objects, on almost any basis it can find, by similarity of shape, of color, of size, of angle, or if the objects are uniform, by proximity. The penchant for grouping is so strong that the mind groups and regroups fields of uniform objects. That is, the mind finds groupings even when there is no perceptual basis for them, to the delight of artists and their audiences. It is not only visual objects that get grouped; sound objects are also grouped, as an essential component of auditory scene analysis and appreciation of music (Bregman 1990). Grouping occurs not just in perception, but also in action; problem solvers use gestures to organize imaginary individuals into imaginary groups (Kessell & Tversky 2005).

Grouping, and other perceptual organizing principles, are processes applied to stimuli in the world. But what about stimuli in the mind, mental objects? Many mental processes, notably, mental rotation, seem to derive from perceptual ones (e. g., Shepard & Podgorny 1978). If the processes occur in the mind, then the objects on which they act can also be mental objects, rather than objects in the world (Finke, Pinker, & Farah 1989). So it is not unreasonable to propose that grouping and other perceptual organizing principles are general processes of the mind, not limited to perception.

Why Group?

Why does the mind group? The most obvious answer is that the mind groups to simplify. The world is nothing if not complex. Grouping objects, putting them in distinct piles, frees the mind to attend to a smaller number of things rather than to each individual. Grouping reduces the number of things that need to be kept in mind. The mind treats the individuals of a group as equivalent, and different from the individuals of other groups. But what of the groups? The mind seems to go yet another step, to assume that groups based on one feature, often a perceptually salient one, share other features, features that may not be perceptual at all. Grouping is effective on one condition: that the feature or features that link the groups do more than that, they also predict or correlate with other features. Perhaps this is part of what is meant by “the whole is greater than the sum of the parts.”

These claims – that grouping is a general mental process and that grouping by one feature is used to predict other, correlated features – need support. To do that, we first consider another enduring contribution of the Gestalt tradition, the study of perceptual illusions.

Errors as clues

From the Gestaltists, psychologists learned that insight into normal perception could come from the study of perception seemingly gone wrong, from compelling and systematic errors of perception. Visual illusions, systematic errors, constrain theories. They also serve as visual jokes, and provide captivating coffee table books. Unlike magic, perceptual errors occur for the simplest stimuli and the simplest judgments. The same line is perceived as longer when oriented vertically than when oriented horizontally, the Top Hat illusion. The same line is perceived as longer when bounded by inward pointing arrowheads than when bounded by outward pointing arrowheads. An oblique straight line looks crooked when interrupted by two parallel vertical lines. Now for grouping, and its’ implications: Two dots that form part of the contour of the same figure are judged closer than two dots, actually closer, that form parts of contours of separate figures (Coren & Girgus 1980). This example also demonstrates the inferences the mind makes about grouped objects, that individuals within the same group are relatively closer than individuals in different groups. Localizing these errors at a specific level of information processing has proved impossible. Surprisingly, the same errors seem to occur at many levels of information processing (Coren & Girgus 1978).

The robustness and ubiquity of perceptual organizing principles suggest that they serve perception. One presumed function of perceptual organization is the establishment of figures, collections of parts that are integral, form wholes. The

test of a whole is another Gestalt organizing principle, common fate. The parts that form an integral whole move together, as a unit (cf. Spelke, Vishton, & von Hofsten 1995). Thus one part predicts another. This is the added value of grouping, one feature predicts another. Might the same organization principles occur in memory, on mental objects? And might the same added value, predictability, result?

Grouping of remembered objects

Telling examples of grouping in memory come from research on cognitive maps (for reviews, see Tversky 1992, 1993, 2005a, 2005b). In a typical study, students are asked to make judgments about directions or distances between pairs of cities. When students in San Diego were asked to indicate the direction between San Diego, California and Reno, Nevada, most incorrectly indicated that Reno was east of San Diego (Stevens & Coupe 1978). Despite the fact that Reno is in Nevada and San Diego in California, neighboring states, Reno is in fact west of San Diego. The likely reason for this error is grouping. Instead of remembering the spatial relations among all cities, people remember the spatial relations among states. They group cities into states, and infer the directions of cities from the locations of states. Since California is generally west of Nevada, people incorrectly make the inference that San Diego, CA is west of Reno, NV.

Other studies suggest that people remember the relative locations of groups as more aligned than they actually are. When asked to judge the direction between Philadelphia and Rome, people incorrectly report that Philadelphia is north of Rome (Tversky 1981). The most likely cause of this error is also grouping. People mentally group the United States with Europe, and remember that Europe and the US are more directionally aligned, even though much of Europe is north of the US, more aligned with Canada, an error termed *alignment*. Although weather may be part of the error, it cannot be entirely responsible as the same grouping or alignment error occurs for North-South directions, where climate is not a factor. People group North and South America, that is, they think that South America is more directly south of North America than it actually is. Thus, they incorrectly report that Boston is east of Rio de Janeiro, when Rio is considerably east of Boston. These effects of perceptual grouping are evident in memory of artificial maps and of blobs, showing that they are broader than geography. Perceptual grouping is also apparent in judgments from perception as well as from memory. Grouping leads most participants to prefer incorrect maps of the world that have been altered in the direction of grouping to correct maps (Tversky 1981). People select a world map in which Europe and Africa have been moved south to be grouped with North and South America to the correct map; they also prefer a map in which

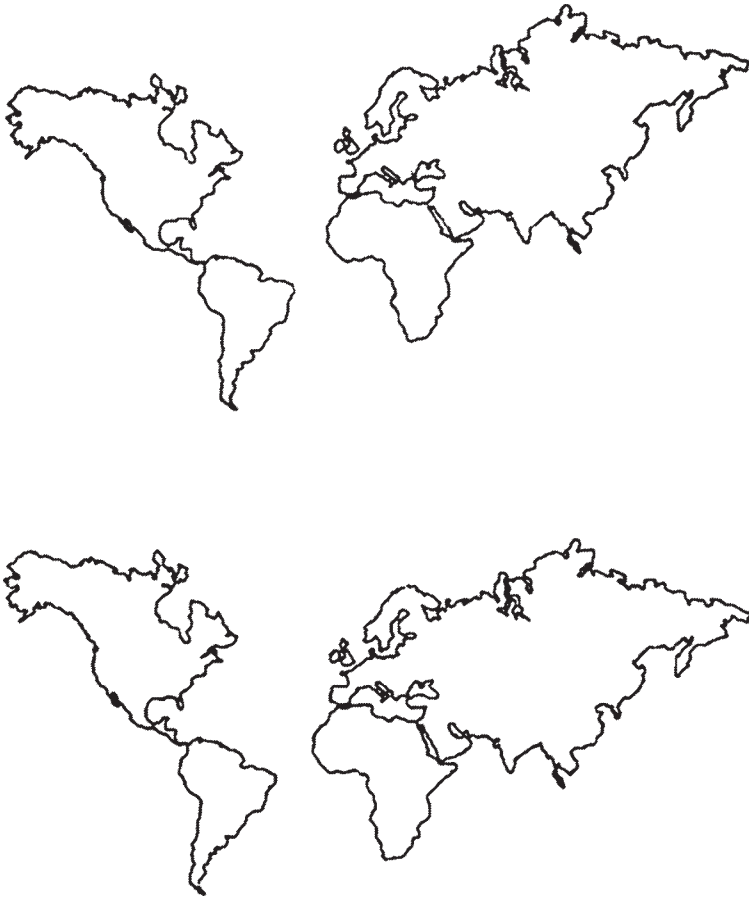


Figure 1. A significant majority of respondents incorrectly chose the lower map over the upper map. In the lower map, the Western hemisphere has been raised relative to the rest of the world so that the United States is more aligned with Europe and South America with Africa (after Tversky 1981).

South America has been brought westward so that it is more grouped with North America to the correct map (Tversky 1981).

Grouping is revealed in times to make spatial judgments as well as errors of judgments of direction. People are faster to judge which of a pair of cities is farther east or farther north when the cities are located in different states or countries than when they are in the same state or country, even if farther (Maki 1981; Wilton 1979).

Grouping by abstract features

These results and others like them establish that grouping can occur for remembered objects just as it occurs for perceived objects. The mind goes yet another step. The mind can create objects it has never perceived, and organize them. What's more, the mind can group imagined objects by conceptual features as well as perceptual ones. Again, the literature on distortions in cognitive maps provides examples. One project had two steps: first, to establish conceptual groupings, a number of residents of Ann Arbor were given the names of a set of buildings in the city and asked to divide them into University or commercial buildings. Next, a second group of residents was asked to estimate distances between pairs of buildings. The distances between buildings in the same conceptual group were underestimated relative to the distances between buildings between conceptual groups (Hirtle & Jonides 1985).

Grouping may underlie landmark errors, in which participants judge an ordinary building to be closer to a landmark than the landmark to the ordinary building (e. g., Sadalla, Burroughs, & Staplin 1980; McNamara & Diwadkar 1997). Cognitively, landmarks define neighborhoods; for example, when asked where they live, most people answer in terms of the closest landmark that they believe

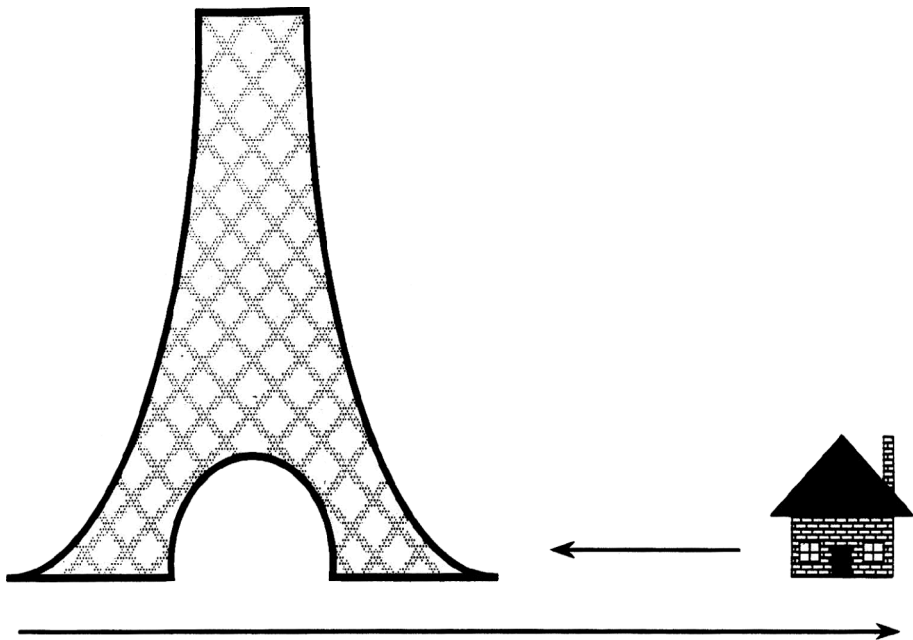


Figure 2. Landmark Effect. Respondents report that the distance from an ordinary building to a landmark is smaller than the distance from a landmark to an ordinary building.

their questioner will know (Shanon 1983). An ordinary building does not define a neighborhood. Thus, when asked the distance from Jacques' house to the Eiffel tower, respondents will group Jacques' house with the Eiffel Tower, in the same neighborhood. However, when asked how far the Eiffel Tower is from Jacques' house, respondents will not group the Eiffel Tower with Jacques' house; they will be in two separate groups, hence perceived to be farther than if they were in the same group. More broadly, landmarks, like prototypes, define or stand for categories (A. Tversky & Gati 1978); ordinary instances just stand for themselves.

Common fate as grouping

Viewed abstractly, common fate can be seen as a corollary of grouping. An entity can be grouped with a large enclosing entity, a reference frame. Once separate things are grouped, they are seen as behaving together, for example, moving together or sharing a common orientation. Common fate biases can also been found in memory as well as perception. Objects induce their own frames of reference. The axis of elongation is viewed as a primary axis and the axis perpendicular to it as a secondary axis, forming an object-centered reference frame. When the reference frame of an object is not aligned with the reference frame of the surround, the two frames of reference are remembered as more closely aligned than they actually

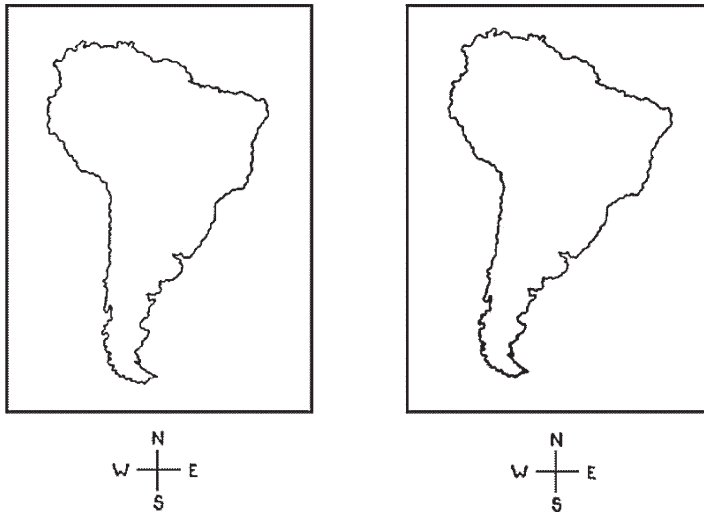


Figure 3. Respondents were asked to place a cutout of South America in a frame. They oriented South America more upright, as in the left-hand part of the figure, than it actually is, as in the right-hand part of the figure (after Tversky 1981).

are, an error termed *rotation* (Tversky 1981). This kind of biased reasoning leads students to upright South America when asked to place it in a NS-EW reference frame. It also leads then to report that Berkeley is east of Stanford when it is west of Stanford, because the Bay Area does not run as north-south as many think. As for alignment, rotation errors occur for artificial maps and cities as well as actual ones, and for memory for blobs as well as geographic entities, implying that it is a general bias.

Grouping in abstract domains

It's time to review. We began with a key contribution of Gestalt psychology, that phenomenologically, perception is organized in ways that cannot be reduced to properties of sensations. Primary among the organization processes is grouping, forming clusters of individuals based on similarity, proximity or other properties. We then speculated that the function of grouping was to simplify, to reduce the number of things that need attention. But to be useful, the grouping needs to be meaningful, that is, the groups need to share properties other than those on which the grouping was based. The mind goes, as Bruner put it, "beyond the information given" (Bruner 1957). Whether or not the groups share other properties, the mind seems to make that leap for perception and memory, or that is what the data reviewed suggest. That inference is what seems to underlie "the whole is greater than the parts."

Levels of gestalt phenomena: From perception to imagination

Phenomenological experience, like perception, occurs at many levels; at each level, experience is related to information given, though it is not reducible to it, ideas developed more fully by Albertazzi (2004). We may perceive amorphous figures on paper as blobs and phenomenologically experience the blobs as grouped. At another level, we interpret the blobs as countries on a map, and phenomenologically experience them as grouped. At yet another level, we retrieve the countries in memory, and experience them as grouped again. These observations about perception hold also for abstract domains (nor are they new, e.g., Bruner 1957; Rosch 1978). The mind can do the same entirely in memory and entirely on abstract concepts. The mind can form groups that never actually appeared together, and form them on abstract features, for example, the scattered members of a political party or a religion or a generation. Although the groupings may be suggested by the features of the stimuli, the mind goes beyond, inferring more features shared within groups and differing between groups. People perceive members of the same group,

social or political, to be more similar than members of different groups, even on attributes not used to form the groups (e.g., Quattrone 1986). At each level, new groupings, new organizations, can emerge. These seemingly inevitable groupings the mind imposes inevitably both allow inferences and entail error; the mind leaps to assume that groups share more features than those that establish the groups and in so doing, the mind ignores differences within groups and exaggerates those between.

Note

* I am grateful to Liliana Albertazzi for stimulating ideas and comments. Preparation of this chapter and/or some of the research reported were supported by the following grants: NSF BNS 8002012 and NSF-REC-0440103, Air Force Office of Scientific Research AFOSR 89-0076, the Edinburgh-Stanford Link through the Center for the Study of Language and Information at Stanford University, and by Office of Naval Research, Grants Number N00014-PP-1-O649, N000140110717 and N000140210534 to Stanford University.

References

- Albertazzi, L. (2004). Stereokinetic shapes and their shadows. *Perception*, 33, 1437–1452.
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.
- Bruner, J. (1957). Going beyond the information given. In H. Gruber (Ed.), *Contemporary Approaches to Cognition*. Cambridge, MA: Harvard University Press.
- Coren, S., & Girgus, J. S. (1980). Principles of perceptual organization and spatial distortion: The gestalt illusions. *Journal of Experimental Psychology: Human Performance and Perception*, 6, 404–412.
- Coren, S., & Girgus, J. S. (1978). *Seeing is Deceiving: The Psychology of Visual Illusions*. Hillsdale, NJ: Erlbaum.
- Finke, R. A., Pinker, S., & Farah, M. J. (1989). Reinterpreting visual patterns in mental imagery. *Cognitive Science*, 12, 51–78.
- Hirtle, S. C., & Jonides, J. (1985). Evidence of hierarchies in cognitive maps. *Memory and Cognition*, 13, 208–217.
- Hochberg, J. (1978). *Perception*. Englewood Cliffs, NJ: Prentice-Hall.
- Kessell, A. M., & Tversky, B. (2005). Gestures for thinking and explaining. *Proceedings of the Cognitive Science Society Meetings*. Mahwah, NJ: Erlbaum.
- Maki, R. H. (1981). Categorization and distance effects with spatial linear orders. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 15–32.
- Malt, B. C., & Smith, E. E. (1983). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior*, 23, 250–269.
- Quattrone, G. A. (1986). On the perception of a groups variability. In S. Worchel & W. Austin (Eds.), *The Psychology of Intergroup Relations* (pp. 25–48). New York: Nelson-Hall.

- Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ.
- Sadalla, E. K., Burroughs, W. J., & Staplin, L. J. (1980). Reference points in spatial cognition. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 516–528.
- Spelke, E. P., Vishton, P. M., & von Hofsten, C. (1995). Object perception, object-directed action, and physical knowledge in infancy. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (pp. 275–340). Cambridge: MIT Press.
- Shepard, R. N., & Podgorny, P. (1978). Cognitive processes that resemble perceptual processes. In W. K. Estes (Ed.), *Handbook of Learning and Cognitive Processes*, Vol. 5 (pp. 189–237). Hillsdale, NJ: Erlbaum.
- Shanon, B. (1983). Answers to where-questions. *Discourse Processes*, 6, 319–352.
- Stevens, A., & Coupe, P. (1978). Distortions in judged spatial relations. *Cognitive Psychology*, 10, 422–437.
- Tversky, A., & Gati, I. (1978). Studies of similarity. In E. Rosch & B. Lloyd (Eds.), *Cognition and Categorization* (pp. 79–98). New York: Wiley.
- Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology*, 13, 407–433.
- Tversky, B. (1992). Distortions in cognitive maps. *Geoforum*, 23, 131–138.
- Tversky, B. (1993). Cognitive maps, cognitive collages, and spatial mental models. In A. U. Frank & I. Campari (Eds.), *Spatial Information Theory: A Theoretical Basis for GIS* (pp. 14–24). Berlin: Springer-Verlag.
- Tversky, B. (2005a). Functional significance of visuospatial representations. In P. Shah & A. Miyake (Eds.), *Handbook of Higher-level Visuospatial Thinking* (pp. 1–34). Cambridge: Cambridge University Press.
- Tversky, B. (2005b). Visuospatial reasoning. In K. Holyoak & R. Morrison (Eds.), *Handbook of Reasoning* (pp. 209–249). Cambridge: Cambridge University Press.
- Wilton, R. N. (1979). Knowledge of spatial relations: The specification of information used in making inferences. *Quarterly Journal of Experimental Psychology*, 31, 133–146.

Visual quality

Drawing on canvas

Liliana Albertazzi

Kandinsky taught the process of seeing, not practical painting.*

Introduction

A critical issue in the analysis of visual space is the difference among *optical space* (generally identified with the space of physics), *phenomenal space* (the anisotropic space of figural and chromatic appearances), and *pictorial space* (as regards both painting and other visual arts like sculpture, ceramics, textiles, computer graphics, and so on) (see Albertazzi, this volume, §1). The difference is not always thematized, for reasons ranging from provenance and disciplinary sectorialization to the tools used for analysis and modelling, and more generally the lack of a systematic categorial scheme of the architecture of spaces.

Numerous areas of contemporary science are still dominated by the Galilean conception of science with its generally tendency to reductionism (on non-Galilean science see Rosen 1985; Ulanowicz 1997). In other words, the *primary qualities* of observed objects still predominate over their *foundational qualities*, so that, for example, the metric qualities of optical space are taken to be the epistemological (and ontological) foundation of every other type of visual quality (see Vishwanath, this volume).

However, psychophysical, perceptual and pictorial analyses have revealed discrepancies in visual space which prevent its reduction to physical space, or the space of position, unless one wishes to consider structures emerging at another level of complexity to be non-veridical or parasitical on previous ones. Current geometries which describe the relations of physical space are therefore inadequate to the task of describing a series of spaces embedded in each other, of differing complexity, and with different exterior/interior relations.

When one admits the hypothetical existence of different spaces governed by different rules of constitution, a number of important questions arise. They concern, for example:

1. The *dependence relations* among the different spaces mentioned.
2. The *role of the perceiver* in these different spaces.
3. The difference between the *unfolding* and the *completion* of visual forms, that is, between the genesis of the form and the objectual quality that constitutes the visual quality of forms as such.

In what follows I shall concentrate on (i) the *phenomenal space of appearances*, (ii) its *genesis*, and (iii) the *forms of subjective completion* involved in the process. I shall conduct a *microgenetic analysis of the development of form* (Sander 1930; Werner, Wapner 1952; Rosenthal 2004), using ‘experimental phenomenology’ as my frame of reference (the definition in Thines 1977).

Because certain laws of the perceptive organization of phenomenal space are similar to those of the organization of pictorial space, the analogy between *many of the subjective processes involved at the primary level* of phenomenal appearances and diverse *modes of drawing on a canvas* is more than being a mere metaphor. That this is so is demonstrated by careful analysis of their structures, of which I shall provide some examples. My treatment will draw on both the experimental literature and the literature on composition in the arts, and specifically pictorial composition in certain avant-garde movements of the twentieth century.

The intrinsic relation between phenomenal space and the space of drawing (pictorial in particular) was discussed by Klee in his diaries, where he writes “the investigation of appearance should not be underestimated; it ought merely be amplified” (Klee 1961:17; see Teuber 1976). In an essay on the “*exacte Versuche*” in art, Klee argued that the analyst should examine, not the complete (i.e. accomplished) form but the process of ‘giving form’ (*Gestaltung*) (Klee 1924:2–3). In many respects, these two statements may serve as mottoes for the analysis of *Gestaltpsychologie*, which was of such importance in Klee’s painting.¹ The theoretical background to my discussion consists of the following:

1. Brentano’s theory of *intentional reference*, in particular the differences among *act*, *object* and *content*, and its subsequent development, theoretical (Husserlian phenomenology) and experimental (Berlin and Graz Gestalt) (Albertazzi 2003; Albertazzi 2005).
2. N. Hartmann’s theory of the *levels of reality*, in particular his notion of the building-above (*Überbauung*) relation (see Albertazzi, this volume §3).
The theoretical writings of Klee and Arnheim (sometimes of Kandinsky) and their ideas on the deployment of visual and pictorial objects, and the relation

with *Gestaltpsychologie* in their works, historically testified by the presence of Duncker and Dürckheim in the annals of the courses given at the Bauhaus.²

3. The microgenetic theory of *Aktualgenese* put forward by Leipzig *Ganzheitspsychologie*, in particular the unfolding of appearances *in phases*, which in many respects resembles Gestalt experimentation at Graz.
4. The gestaltist *scientific literature*, both old and recent.

Given that the theory underpinning *Gestaltpsychologie* is almost entirely forgotten and that Graz Gestalt theory is almost entirely unknown, and that their vocabulary and conceptions are almost entirely alien to the contemporary theory of representation and epistemology, I shall begin by discussing them briefly.

Descriptive analysis of seeing

Brentano's theory comprises an idea of *geometric imagination* which springs from Herbart's concept of *serial forms* (*Reihenformen*) according to which, for example, the presentation of something spatial and/or temporal requires a succession *in* the act of presentation (Herbart 1816, Part 3, Chapter 1; Albertazzi 2002).

Brentano defines as an *act of intentional reference* every presentation (*Vorstellung*) that we acquire through our sense perception *or* imagination (the term *Vorstellung* in German has the characteristic of being *internal* and tied to *imagination*, see below). Examples are *hearing* a sound, *seeing* a coloured object or *feeling* warm or cold, as well as similar images in imagination (*phantasia*).³ Further examples are *thinking* a general concept, and every *judging*, *recollecting*, *expecting*, *inferring*, *persuading*, and *feeling* (Brentano 1995: 88, 79). The use of the gerund emphasises the act's nature as a *process* rather than as a state.

The *objects* of presentations are, for example, *a colour*, *a figure*, *a landscape* which I see, *a chord* which I hear, and then *warmth*, *cold*, *odour* which I sense, and the corresponding images that appear in imagination (Brentano 1995: 79–80). The noun form indicates their nature as correlates, or in Brentanian terms, as non-independent parts of the act in its deployment (Brentano 1995). In other words, as the act comes about, it is impossible to separate it from the object at which it is directed. It is a unitary process which takes place rapidly, lasting some dozens of milliseconds to one or two minutes. Consequently, descriptive analysis of its parts seems to be only possible when the process has been completed.

Detailed analysis of presentations, from both the theoretical and experimental points of view, leads to their configuration as concrete spatio-temporal structures ('things' that actually occur), as subjectively integrated processes, and as processes of primary type (for example, seeing a colour, a shape or a landscape). As Klee put the same idea: "The object is surely non-existent. It is the sensation of the object

that is of first importance”, by which he meant the inner construction of the form (Klee 1961: 191).

Experimental studies on temporal apprehension have shown that the dynamic of the time of presentness (in technical terms *psychische Präsenzzeit*, henceforth TP) (Stern 1897, 1898; Benussi 1913; Bonaventura 1929; Katz 1906; Wirth 1934) possesses very fine granularity. In particular, experimental analysis has revealed that the temporal units of the perception of forms are not punctiform instants but have an extension (*Extensität*) (Brentano 1988; Albertazzi 2002, 2005, Chapter 7) which may vary according to layout, content, attention, and so on. The units of time that constitute an act of presentation are therefore not of metric type, they are not instants: they are moments (*Momente*) endowed with places (*Lage*) within the duration which may stretch or shrink, and their nature is essentially *qualitative*. However, the qualitative nature of the temporal units does not prevent their experimental psychophysical analysis (Fraisse 1964, 1974; Michon 1967; Michon, Jackson 1985; Libet et al. 1979; Libet 1982). The problem is instead identifying the dependence relations between an external psychophysics (based on atomic, metric units of representation) and an internal psychophysics (based on the other type of representative unit, non-atomic and non-metric). In other words, this is a matter of bringing Fechner’s original project to completion.

The works of Klee, Kandinsky and Picasso were driven by the same intent to analyse the categories of internal and external at the level of visual perception in the aesthetic field. As the Cubists proclaimed, there is nothing real except the coincidence of a sensation with an individual mental direction. Which does not mean to cast doubt on the existence of the stimuli that strike our senses. However, we can only have certitude with regard to the images that they produce in our mind.

As for the objects of a visual presentation, Brentano pointed out that they may be of differing complexities; they may be, that is, qualities, figures, scenes, or images of the phantasy. Whatever the case may be, they are objects of some type given in the actual presentation, in a particular here and now (*Jetzt*).

There are essentially two types of presentations:

1. *Perceptual* presentations (seeing, hearing, touching – a present object).
2. *Mental* presentations (thinking, imagining, remembering – an absent object).

This terminology, having passed through the developments made to *Gestalt* by the Graz School, is current in the contemporary theory of vision, where it takes the form of the better known Kanizsa’s distinction between *perceptive presence* and *mental presence* (Kanizsa 1991, Chapter 1).

The role of the perceiver, given that s/he is the *bearer* of the actual presentation, forms an integral part of the structure. It is therefore a functional component, not an Ego, nor a self in the generally understood sense of ‘subject’. In the brief elapse of the TP, in fact, the predicative distinction between subject and object is literally

impossible, because it is the *antepredicative structures* of experience, as they are being produced, which are under observation (Husserl 1966b). This also explains why the 'objects' of the seeing process are necessarily integrated subjectively.

The existence of this *subjective structural component* of the presentation makes also the space of the presentation proprio-centric. This is shown by the role performed by the various subjective integrations in perception, or by the fact that data-driven information (the standard contemporary concept of bottom-up information) is modified in the presentative content of the time of presence. Consider, for example, the objectual complexity of a melody in acoustic perception, which is still recognizable even when all the notes have been changed, or the perceived three-dimensional form of a room, which does not change even when we shift our point of observation, and therefore despite the fact that everything changes at the level of retinal stimulation (Ehrenfels 1890; Meinong 1899).

The bearer of the presentation, therefore, should not be understood in the sense of an 'ideal presenter', and even less in that of the 'ideal observer' of contemporary epistemology from physics to linguistics. Finally, the essential anisotropy of visual space requires us to assume, like Benussi, that *in principle* every presentation is intrinsically multiple.

Another fundamental feature of the presentation is that it *presents or exhibits the structure of what appears* in the TP. Strictly speaking, therefore, it is *not a representation of the 'external' world* understood as the source of stimuli (see Introduction, §2). This roughly corresponds to the concept of 'bracketing' (*epoché*) that characterizes phenomenological inquiry (Husserl 1913, I, §§31–32). No phenomenologist either theoretical or experimental, from Husserl to Michotte, has ever doubted the existence of stimuli with a triggering effect on the structures of the presentation, although they point out that this is not the only type of causality that exists. Much more important for this type of analysis is the structure of inner causality, which leads to identification of an object via the temporal (and kinetic) structures of the presentation (see below) (Husserl 1966a, where the modes are retention and anticipation; Leyton 1992).

Once again, therefore, the specific interest of a phenomenal theory of appearances is not as generally understood in the contemporary theory of vision (a survey in Vishwanath 2005), namely:

1. The detection of a stimulus array.
2. The transformation and integration of multiple primitive components, viewpoint invariant.

Bracketed off are both the stimuli and the so-called transphenomenal correlates of consciousness – or at least they are so at this initial level of analysis (Husserl 1913, §§31–32). The next step in description of appearances in all their complexity is analysis of the relationships among stimuli, transphenomenal correlates, and

appearances. However, the procedure suggested by phenomenological analysis – starting from the description of appearances – may give more correct focus to research by psychophysicists and brain scientists themselves.

‘Presentations’ (*Vorstellungen*) are the *structures underlying phenomenal appearances*, or the *system of relations* that govern them. Analysis has shown that these are *anticipative* structures which:

1. Unfold processually,
2. Are endowed with direction,
3. Give rise to the perception of continuity (often also amodal).

These are *deployment acts* which extend through a continuous series of durations comprising ‘fringes’ of both the past and the future.⁴ As said, the ‘atoms’ of inner time are not instants, and the temporal relation that governs the unfolding of presentations is not of before/after type, as in metric temporality, but one of *simultaneity in succession* (Albertazzi 2005, Chapter 4).

Having performed this bracketing operation, which highlights the intrinsic temporality of every type of perceptive presentation, *descriptive* analysis of the objects of seeing identifies its constituent components as *colour, form* and *position in visual space*.⁵

Also to be borne in mind are the following facts:

1. We experience *a huge variety of objects* of visual presentations: cars, people, movements and (so-called) apparent movements, blots, pictures, statues, and so on, of increasing complexity (Albertazzi 2003).
2. The objects of seeing are *never simple* (because they are given in a twofold figure/ground representation, or with masking, occlusion, anomalous contours, etc.). They are therefore Gestalt wholes given in a spatio-temporal coherence field.
3. Visual space is a *potential space* because it is the locus of the presentation of appearances located in simultaneously present places ordered temporally and spatially (Husserl 1966a, 1966b; Koenderink 1990).

In short, the ‘objects’ of visual presentations in TP are:

1. *Complex* appearances with differing degrees of granularity (say, from ‘proto-objects’ to objects) (Rensink 2000).
2. *Internal* correlates of seeing, i.e. non-independent parts of the act.
3. Endowed with *evidence* because they are *co-present* to the act of seeing itself.

From this point of view, analysis of presentations does *not* concern itself with the question of *epistemological veridicality*, because it is only meaningful to enquire as to the veridicality or otherwise of so-called multistable figures – to cite a well-known example – if the physical level of stimuli is taken to be their primary level.

But the majority of the objects of vision operate on the *presentational* unfolding structures of vision, or are *field objects*, i.e. they have another primary level of reference. It is in this that consist both the concepts of 'intentional reference' and of 'bracketing' in descriptive analyses of phenomenal appearances. All that phenomenology does, therefore, is recognize the (relative) independence of an emergent qualitative level of reality from one or more underlying physical levels (Ehrenfels 1890; Koenderink, van Doorn 2003: 251; Albertazzi 2005), while also, obviously, accepting that this level must be analysed *iuxta propria principia*.

Given these premises, in what does a theory of visual appearances consist?

Act versus object

Klee distinguishes between act and object in the light of the difference between linear acts and lines, or in other words, between the visible and the invisible dynamic structures underlying it and which 'produce' it (Klee 1961: 112).

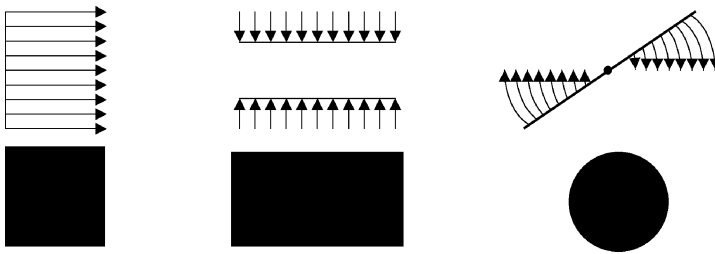


Figure 1.

In actual drawing, a square results from the translation of forces in the same direction, while a rectangle becomes such through the translation and compression of the same forces in opposite directions. A sphere is the result of rotation of the drawing around a point.

The distinction between visual (visible) properties and hidden, but active and appearance-producing, structures in the objects of vision has also been analysed by Arnheim (Arnheim 1954, Chapter 1).

In this pattern, for example, the asymmetric position of an eccentric disc in a square is due to the *inner relational properties* (i.e. tension) of the whole. The disc is seen as occupying an 'eccentric' position neither because of previous experience nor because of some sort of logical inference. The centre's power of attraction, in fact, is intrinsic to the *inner relations* of visual patterns because it is embedded in a field of forces, in an anisotropic and temporally dynamic space – as shown by analysis of the '*structural skeleton*' of form.⁶

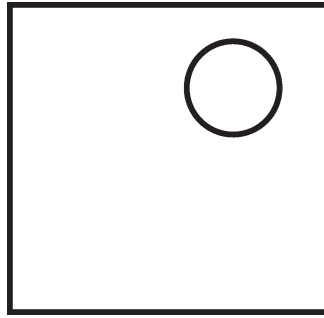


Figure 2.

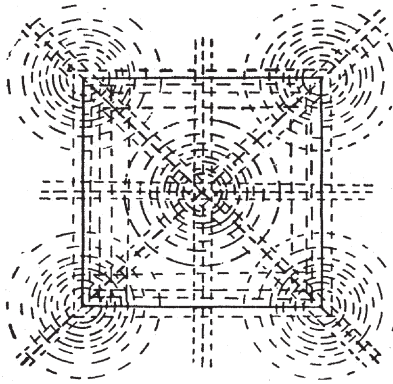


Figure 3.

The structural skeleton of the square in this case is determined by the cross formed by the vertical axis, the horizontal axis, and the diagonals, in a dynamic relation.

Itten has analysed the nature of ‘good’ and ‘bad’ points in patterns of this kind, identifying the intrinsic relation between the positions of the parts with respect to the whole and tertiary qualities (Itten 2002: 72–73).

Another interesting aspect is the relation between the structural skeleton and the *orientation* of the form. If a triangle or a rectangle is inclined, it remains a triangle or a rectangle. If we rotate a square, it becomes a rhombus or a diamond, but in this case the structural skeleton do not shift with the figure. A new symmetry makes the principal horizontal and vertical axes pass through the angles, transforming the sides into oblique shapes. In this new mode of appearance, the diamond is a new figure, more dynamic and less anchored.

The underlying intrinsic mobility in the unfolding of the visual field is made explicit by the action itself of drawing, which develops following the opposition of

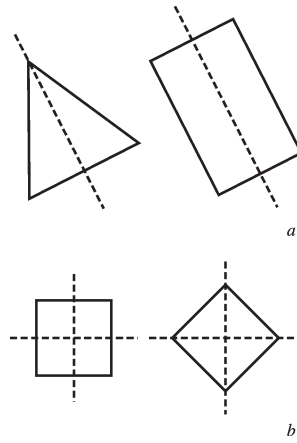


Figure 4.

movement and counter-movement from one overlapping plane to another. How an appearance is produced in movement, starting from a generator centre until a maximum extension constituting its boundary is reached, is once again illustrated by Klee with the following example (Klee 1961: 16).

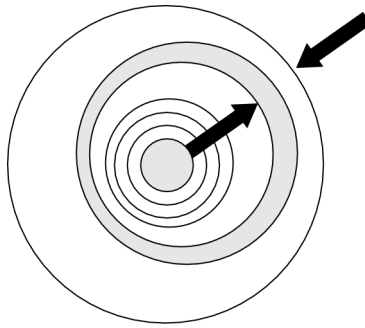


Figure 5.

This microgenetic deployment shows the outer/inner, figure/ground, before/behind, left/right oppositions implicit in its development as *modes of deployment* (i.e. appearing) of the *structure*. Thus the visual object constructed within the structures of the presentation is given by the *outer boundary of the inner constructive form*. It is for this reason that phenomenology conceives the meaning – and the value – of an object in the contentual structure derived from the interplay among the inner relations of a pattern (Meinong 1899) as the coherence of a dynamically complex whole. Klee makes the same point when he remarks that when a shape is drawn, the “whole inward territory is designated by the word *content*” (Klee 1961: 127).

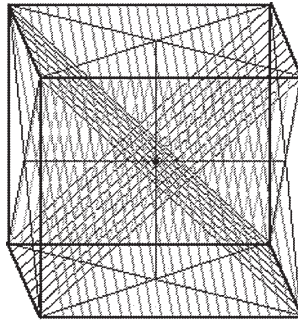


Figure 6.

The geometry of unfolding appearances

That even apparently simple ‘visual objects’ like a cube, a chair or a lemon, which appear to be static and devoid of inner movement, are nevertheless the product of forces and tensions – that is, are the product of a process – can be demonstrated by a variety of examples. Those now described are again taken from artistic theories and the experimental literature.

The concept of *Gestaltung of direction* is evidenced by Kenkel’s γ movement. The experiment is as follows.

The tachistoscopic exposure of a cardboard disc for ca 100 ms shows the expanding movement of the internal part of the figure (Kenkel 1913).

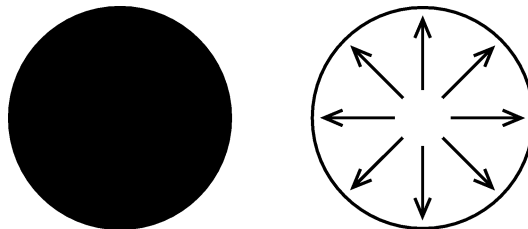


Figure 7.

The γ movement is an excellent example of the unfolding of the presentation because it evidences various features of the process:

1. The focal point of γ movement has *central symmetry* in all directions.
2. The object appears to grow, from the point of fixation, as an *internal process of expansion* (i.e. it is not something with stable properties but an appearance that deploys).
3. The γ movement evidences the deep lying structure of appearances as the *movement of the inner parts* of an appearance as a pull in several directions.

This pull, however, it is not the same for all parts: the horizontal prevails. Moreover, the prevalence changes according to whether a square, and oval or a triangle is being presented.

4. The γ movement evidences the *distortion of the margins* of brief processes.
5. The γ movement is a phenomenal expression of the *arousal* and *disappearance* of visual forms (it is, moreover, that we frequently encounter, for example when we light up a dark room and the objects seem to expand).
6. The final form tends lawfully towards stability, according to the conception of *Prägnanz*.⁷

Whilst the γ movement highlights the *development of the inner parts* of an appearance in the TP and the *behaviour of the direction* of deployment, another movement analysed by experimental phenomenology exhibits the relation between object (appearance) and content in the development of a form, and the role played therein by direction. These are the stereokinetic movements analysed by Benussi and Musatti from the 1920s onwards (Musatti 1924; Albertazzi 2004).

One of the most interesting results of Musatti's experiments concerns stereokinetic phenomena with the rotation of *curvilinear figural complexes with variable curvature*, in particular the rotation of black cardboard disks with curved white lines drawn on them (Musatti 1955, §3, 13–19).

Suppose that the drawing is an ellipse. In this case, the movement splits into two movements, one along the ellipse, which is not seen, and another that moves in a radial direction. The figure no longer rotates (the rotational movement disappears), but all its points move radially towards the perimeter or towards the centre. Apparently, therefore, every point of the ellipse maintains its direction towards the centre of the disk, only changing its distance.

In this case, the *stability of orientation* is at the expense of the figure's rigidity. What one sees is an elastic ring which changes shape because each of its points is constantly in movement with respect to the others (it 'writhes').

The stability of orientation is not perfect. It stretches for each element in the figure in a manner difficult to describe because it changes according to the point of observation.

Together,

1. Stability of orientation and
2. Continuous deformation

give rise to a *libration of the figure in space* (stereokinetic transformation) which also loses the property of deformability. One now sees a disk or a rigid ring which librates in three-dimensional space, constantly oblique and oscillating without rotating.

In the case of the ellipse that distorts into a ring, the tendency to the minimum trajectory holds. In the case of the ring or disk positioned obliquely in three-dimensional space, the tendency to constancy holds. Thus we have:

1. Rotating ellipse, constant shape, on a single plane.
2. An elastic ring or self-distorting figure on the plane and stably oriented.
3. A disk or constant figure, stably oriented and shifted in space (Musatti 1955: 16–19; Zanforlin 2000).



Figure 8.

In view of these findings my hypothesis is as follows:
The phenomenal space of appearances is given by:

1. A generating centre and its expansions.
2. A tension between form and function.

As we have seen, this is a conception in many respects similar to those of Brentano, Klee, Kenkel and Arnheim. As for Arnheim in particular, my hypothesis resumes the already-discussed concept of the structural skeleton underlying the perception of form (Arnheim 1954: 91). The structural skeleton of a form, in fact, is given by:

1. The scheme of the *axes of direction*.
2. The corresponding characteristics of the *parts* created by the axes.

It is not given by its *contour*! The relation between the structural skeleton of a form and the completed form (i.e., the final visual ‘object’) is rather complex, because the *same structural skeleton can be covered by a wide variety of forms*. Then, if the visual pattern can give rise to diverse structural forms, it can also be perceived, for example, as two different ‘objects’ owing to the presence of two contradictory structural skeletons, in opposite directions. This situation has obviously important consequences for modelling of the phenomenon too.

One of the best examples of the existence of competing phenomenal directions underlying an appearance is provided by the so-called Jastrow illusion.

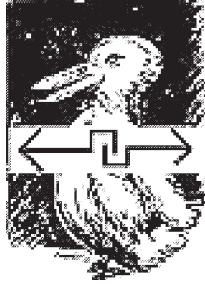


Figure 9.

The two spatial zones that determine the perception are characterized by different directions: towards the left (duck) and towards the right (rabbit).

Specifically, the thesis put forward by Arnheim in *Visual Thinking* is as follows. Perception consists in the *adapting* of stimulus material to relatively simple form moulds (patterns, types), or generic structural components of the percept, like 'roundness', 'canineness', 'straightness', 'lineness', etc. (Arnheim 1969, Chapter 2; see also Ingarden 1962, Chapter 3; Husserl 1966b).

Given these premises, what kinds of qualities do appearances possess?

Qualities

The classical partition distinguishes between primary and secondary qualities. *Primary* or Galilean qualities are shape, extension, number, motion, according to the laws and definitions of classical physics. They are numeric and quantitative. *Secondary* qualities, a product of British empiricism, are qualitatively connoted, and they are for example colour, odour and taste.

The Berliners distinguished a third type of qualities of forms which possessed an *Afforderungscharakter*, i.e. a demand character, a valence. These types of *tertiary qualities* pertain to the *experienced* object, and today are termed 'affordances' in Gibsonian ecological theory. The original Gestalt conception, however, was much more complex (Albertazzi 1997).

On the basis of these distinctions and the foregoing discussion, I shall now recast the theory of the visual qualities of appearances, stressing the following points.

1. Visual qualities are not physical qualities like metric qualities.

2. The morphogenetic order of the qualities does not correspond to the classification of physical qualities.
3. Various kinds of subjective integrations are inner determinants of a specific quality.

I cannot claim that my proposal is entirely new, given that other authors of various provenance have put forward similar ideas. Lipps, for example, defined the visual qualities as *distinguishing features*, thereby contrasting them to Fechner's notion of *distinctive* (i.e. elementary) *features* (Lipps 1897; Albertazzi 1998). Husserl called them *moments* or non-separable parts of the qualitative wholes of appearances (see above) (Husserl 1966b; Arnheim 1969, Chapter 2; Ingarden 1962, Chapter 3). Both these authors, therefore, emphasised the qualitative and intrinsically meaningful nature of visual forms.

Wertheimer, Kandinsky and Arnheim termed them *patterns* and stressed their dynamic structure. The pattern is the dynamic component which regulates the law of similarity in the perceptive organization. This law states that similar components tend to be perceived as pertaining to the same unit, and it holds in both visual organization and other sensory modalities; for example, it also holds in auditory organization as regards the contrast between loud and soft sounds.

My proposal is that the *primary qualities* of appearances as they constitute themselves in the TP are aspects such as 'squareness', incomplete or potential, 'angleness' which constructs the square, or 'straightness' as vertical/horizontal segment, or 'curvedness', 'parallelness', 'lineness', 'acuteness', 'thinness'. These qualities *persist in the perceived form* even after its completion as *traces of the dynamic patterns* underlying their constitution.

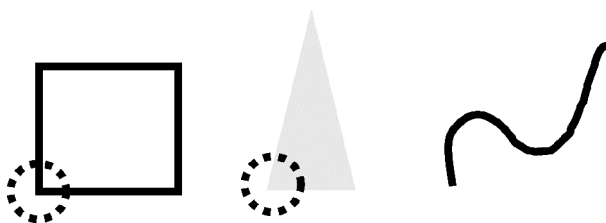


Figure 10.

I consequently define as secondary qualities, for example, both phenomenal blue *and* a particular nuance or shade of blue, while tertiary qualities are dynamic, diffuse, expressive properties. Examples are 'round', 'angled', 'sharp', 'uncinate'. As *such*, these qualities are also 'elegant', 'common', 'noble', fearsome', etc., in that they acquire aesthetic and value meaning in both visual and pictorial 'objects'. There are abundant examples of these invariants transposed into pictorial objects. As

Klee noted, in fact, in perfect consonance with Wertheimer's theory: "Every form has its face, its physiognomy. Shapes look at us, gay or severe, tense or relaxed, suffering or smiling" (Klee 1924).

In short, I argue that in the visual space of appearances what are generally called primary qualities (conventionally understood as metric) *must be bracketed off*, because at the level of appearances the 'object' is non-transcendent. Specifically, the primary qualities of appearances:

1. Are not symbolic.
2. Are not atomic.
3. Are not indexes of external entities.
4. Are not viewpoint invariant.

Above all, they are *qualities presenting themselves*. For example, in the following pattern:

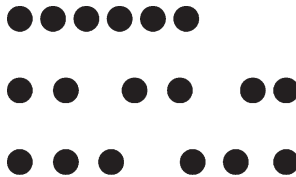


Figure 11.

we have neither a set of points nor, strictly speaking, three rows of points. What we instead have, according to the laws of organizations of percept (Wertheimer 1923), is:

1. A row of points.
2. A row of pairs of points.
3. A row of triplets of points.

Above all, they are *not* signs of something else!

Although we do not have a complete theory of visual qualities from a phenomenal point of view, we at least have an initial classification of them drawn up by Metzger (Metzger 1941, Chapter 2), who distinguishes among:

1. Qualities pertaining to the *tectonic* structure, like spatial form, configuration, distribution of colour. And, among these, a subset formed by those specific to the *dynamic structures*, like 'direction', 'tension', 'change', 'attraction', 'push', 'launch'. Various qualities of this kind were discussed above. Qualities of the 'push' and 'launch' type are manifest, for example, in the phenomena of causality perception described by Michotte.

2. *Global qualities* of the phenomenal ‘material’, for instance ‘transparent’, ‘bright’, ‘corporeal’, ‘incorporeal’ in reference to shadow and light. Stereokinetic phenomena display properties of this kind, and so too do the shadows or the brilliant white of Kanizsa’s triangle.
3. *Expressive* or properly tertiary, qualities having physiognomy, valence, demand character, or tonus, as already analysed.

My particular contribution to Metzger’s initial outline of qualities is (i) to stress their *inner order of morphogenesis*, (ii) their *microgenetic structure* in TP, and (iii) emphasising in all three cases the role of the subjective integrations as ‘inner determinants’ of the specific quality. Consider, by way of example, how the attention can change the strength of a stimulus by increasing its ‘effective contrast’ or salience (Carrasco, Ling, & Read 2004).

Inner determinants

From the point of view of actual seeing, the tertiary qualities are *morphogenetically* primary, while the structural ones are *methodologically* primary. This difference accounts for large part of the diversity of analysis conducted, within *Gestaltpsychologie*, between the experimentalists of the Berlin School and those of the Graz School. In other words, it accounts theoretically for the difference between analysis of the act and analysis of the content/object in vision.

These different types of analysis also produce different results on the basis the same kind of experimentation. For example, analysis of the inner structure of the act of seeing explains the apparent magnitude of the parts of the following visual pattern (made famous by Lipps) (Lipps 1897; Klee 1961: 127; Meinong 1899; Husserl 1900–1901):



Figure 12.

Benussi sought to identify the *subjective structures* that are in part responsible for this specific appearance: namely, the *phenomena of assimilation*. He found that the apparent size of the parts of the pattern in anisotropic visual space derives from the *subjective mode of comparing the parts*, which cooperates to the homogenization of the whole. This *subjective operation* is made possible by the structure of the

TP, which deploys in a simultaneity of parts in succession; or, in other words, by the existence of *phases within* the TP.

Another subjective operation that takes place within the TP is the mode of *connecting* the parts of a pattern, which are highly visible in phenomena of multi-stability (Benussi 1906; Meinong 1904), for example in the following pattern:

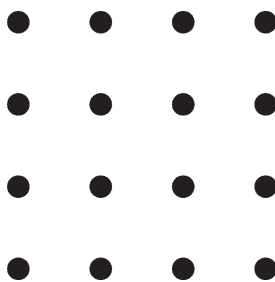


Figure 13.

These *functional parts of the structure of the act* (the ‘inner determinants’ of the act in Benussi’s old-fashioned terminology) constitute only one aspect of the complex architecture of the *seeing* which deploys in visual space. For example, it is because of the complexity of the microgenesis of the structure of appearances that visual space is immediately presented with ‘negative’ or ‘positive’ aspects, exactly as in the above-mentioned case of points. In fact, a concave contour is seen as the edge of a whole, while a convex contour is seen as a figure; however, reversal is given by forwarding cuneal wedges, because of the diverse distribution of attention on the different parts (Arnheim 1954: 203). This complexity, which rules the dynamic between function and form in visual appearances, is due to a specific kind of *inner causality* which regulates its unfolding.

That *perception is not acted upon by the stimulus* has been demonstrated by the phenomenological approach to the experimental analysis of vision (see above). Moreover, there are various ways to analyse its complexity, ranging from Husserl’s studies on inner time consciousness (by the way, the subject of one of Varela’s last essays) to Benussi’s experiments on acoustic perception as described in his *Zeitauffassung* of 1913 (Husserl 1966a; Benussi 1913). Evidence for the existence of this structure in the visual field is provided by two examples from Massironi, who illustrates two presentations of causal relationship internal to the visual field: a texture deformed by a diamond and a texture deformed by a rectangle (Massironi 2002: 207).

These two examples, despite their apparent simplicity, highlight various aspects of the complex architecture of the act of seeing and its phenomenal output. They show, for example:

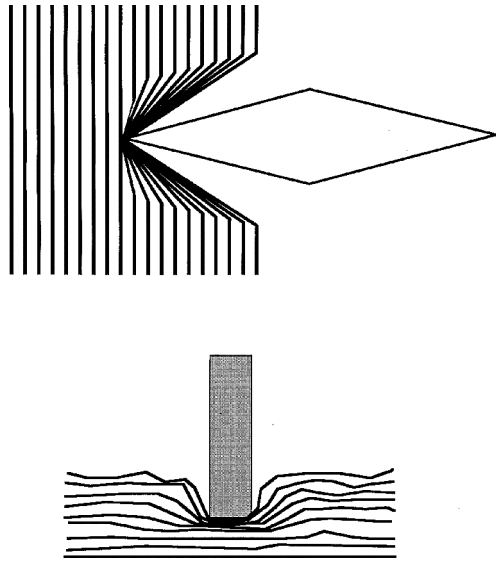


Figure 14.

1. The *anticipatory structure* of the TP, the unfolding of which comprises ‘fringes’ of the past and of the future which are retained and reactivated by the apparently static visual pattern.
2. The *expressiveness* of the lines (the rhombus has the ‘sharpness’ of a cutting implement, the rectangle has the ‘weight’ of a brick).
3. The process by which the *identity* of phenomenal objects is constituted (Massironi 2002: 209).

That *time is a cause* in the sequence of visual appearances, and that the perception of causality in a static image is due to subjective processes which construct the inner continuity of succession into a causal chain, are also tenets of the theory developed by Leyton. For Leyton, in fact, the *history of a shape* is immediately visible in the lines of its configuration. Specifically, he claims that the breakdown of symmetry is at the basis of the conveying of temporal information through images, and that a single *abstract* property characterizes *all* perceptual situation of memory (Leyton 1992: 7). As Massironi notes, however:

The critical feature is not asymmetry per se, but the fact that asymmetry is a discontinuity relative to an otherwise symmetrical structure. More generally, one could speak of a trade-off between regularity and irregularity in the comparison between local parts of a pattern and its global structure. This approach demands that our phenomenon should be possible also in a globally asymmetrical structure possessing a local symmetry. (Massironi 2002: 209)

This observation also highlights the fundamental characteristic of my proposal to consider the microgenesis of form of appearances in phases, whose deployment consists in the unfolding of the whole in the inner relations that arise among the parts in the TP. The microgenetic approach therefore concerns itself with the *prehistory of shape* and its acquisition of visibility through the ‘deformation’ and reconstruction of the stimuli. Psychophysical analysis in the TP, in fact, seeks to make the genesis of things visible.

To sum up, the difference between form and function in visual appearances consists in the following:

1. The unfolding structures of TP.
2. The presence of an invisible structural skeleton.
3. The various types of quality.
4. The part/whole relations.
5. The inner causal dynamic.

I said at the outset that one of the main issues concerning visual space concerns the difference among optical space, the phenomenal space of appearances, and pictorial space.

Given that, as I have sought to show, the phenomenal level of appearances is structured in terms of dynamic patterns – beneath which specific unfolding structures are responsible for their *presentation* – the next question is as follows: How can these patterns be *represented* in their complex interdependency? The question is important from both an epistemological point of view – given that it is at odds with the standard theory of representation in the cognitive sciences – and from the methodological point of view of modelling. My view is that analysis of the relationship between phenomenal space and pictorial space yields insights into how the problem should be framed (see also Hecht, Schwartz, & Atherton 2003). On the basis of the phenomena already analysed, I shall now furnish an example of such analysis, doing so by concentrating on just *one* aspect of the problem, that of *deformation* in visual patterns.

Rule-based deformation

As I have repeatedly pointed out, in phenomenological analysis the category ‘stimulus’ fulfils the role of the *Restbegriff* or limiting concept, which is certainly not denied but is considered pertinent to another domain of ontological reference. A stimulus, in fact, once it has been elaborated neuronally, undergoes further *modification/deformation* by the temporal and spatial structures of the TP and assumes qualitative properties. It is only by an operation of causal inversion that the stimulus is made to coincide with the percept ‘seat’, for example. Vice versa, the

percept 'seat' is the result of a complex series of visual, motorial, tactile, functional, aesthetic and conceptual operations.

That *deformation* is a key factor in perception, and especially in the perception of depth, is well known. Deformation manifests a change in spatial structure and a distortion in the canonical form. We have seen striking examples in stereokinetic phenomena: for instance when an oval pattern on a spinning piece of black cardboard gives rise to a temporary perception of distortion which subsequently stabilizes into a three-dimensional appearance (a coin or a cylinder, depending on the initial pattern). As I have pointed out, tied to this aspect of the perceptive structure of appearances are the qualities which Metzger and Arnheim classified as dynamic: 'pushing', 'pulling', 'compressing', 'folding', 'bending', etc. (Arnheim 1969, Chapter 5). Extreme cases of this fundamental feature, which once again demonstrate the *inner causal structure* of appearances, are anamorphic forms for example.

Like all the products of the inner constructive form – i.e. the interplay between structural skeleton, causal structures, phases of deployment, and the part/whole relation in visual patterns – this type of deformation has *presentative character*: in other words, it only 'presents its own unfolding'. This, therefore, is not a *representation* of something external to the pattern, but rather an *expression* of the pattern which retains and gives form to the latter's constitutive invariants.⁸

The genesis of patterns in the subjective space of inner presentation is evidenced by the role that it performs in pictorial space, which expresses it and magnifies its specific patterns.

The classic loci of deformations in pictorial space are decoration (arabesques, stained glass, pottery) and modern and abstract art, specifically painting, where, for example, the question is *how to compose an internally closed whole* with colour patches (Ingarden 1962, Chapter 9).

As Klee writes:

To be an abstract painter does not mean to abstract from naturally occurring opportunities for comparison, but, quite apart from such opportunities, *to distil pure pictorial relations*: light to dark, colour to light and dark, colour to colour, long to short, broad to narrow, sharp to dull, left-right, above-below, behind-in front, circle to square to triangle. (Klee 1924: 72. Emphasis mine)

We have already seen an example of the genesis of these pure pictorial relations when analysing the dynamic between movement and counter-movement in drawing: this process shows the genesis of forms below the surface.

My working hypothesis – within the framework of the theory of intentional reference and its deployment in the act of seeing – is as follows:

Abstract painting shows the structure of the phenomenal appearances in that it is a case of reverse intentionality.

In other words, the pictorial form brings with it *traces* of the structures underlying the *construction* of visual forms and, depending on the style, displays certain fundamental aspects of it. I therefore adopt Arnheim's idea that pictures provide *graphic equivalents* for objects and scenes. This point requires specification.

For example, in this volume Willats notes that, because of the development of computer vision, it is now possible to describe these graphic equivalents in terms of *three* representational systems:

1. The drawing systems that map spatial (topological) relations in the scene into spatial (topological) relations in the picture.
2. The denotation systems that map scene primitives (volumes) into picture primitives (regions).
3. The mark systems (lines) that represent picture primitives by physical marks (outlines).

However, my analysis here is not yet concerned with the problem of the reproduction and simulation of seeing with machines; at the moment I am dealing with the relation between *actual seeing* (presenting) and *representing through actual painting*, my purpose being to show their shared laws of visual organization. In my vocabulary, then, the term *primitives* means the *semiotic patterns* created by painting which, as non-independent parts, acquire value only within the figural composition of the picture.

The parts perform a specific (dynamic) role in the configuration of the whole, as the Berliners stressed (*haben die Teile eine Rolle!*) (Kopferman 1930). Suffice it to consider the diverse parts that are obtained from a two-dimensional or three-dimensional configuration of a cube and/or their intermediate forms. As Husserl pointed out, moreover, the configurational whole of a visual appearance (a cube, for example) does not consist summatively of a set of independent parts (pieces) or of viewpoint-invariant features. On the contrary, the whole (be it a cube or a tree) consists of the constant modification and deformation of the parts, which are also apprehended amodally (for example, its rear) in simultaneous continuity from a specific point of view. In Husserlian phenomenology, these diverse 'views' of the object are denoted with the technical term *Abschattungen* (adumbrations).

To what extent do the parts of a *physical* object differ from the parts of a *visual* object? For example, inspection of how the preparatory drawings for construction of the Eiffel Tower affords interesting insights into the various representations involved. On the one hand, the drawings are graphic representations in which the hierarchy of space is presented as a function of distance, and the object is depicted from a particular point of view. This type of representation closely resembles natural perception based on non-independent parts. On the other, the drawings are representations of disassembled parts (pieces) which illustrate the various engi-

neering operations required to construct the object (De Vries 1971; see Massironi 2002: 92–96).

Delaunay produced a pictorial depiction of the Eiffel Tower which uses deformation to convey the *non-summative unity of the visual whole* in the simultaneous continuity of the various viewpoints and of successive looks at it. In other words, Delaunay's *Eiffel Tower puts into form, expresses*, and in this sense '*represents*', the complexity of an act of phenomenal representation and conveys its intrinsic causal formation. From this point of view, Delaunay's *Eiffel Tower* resembles the first type of depiction made by the preparatory drawings of the Tower, but is made more complex by the inclusion of all possible viewpoints of the object and their dynamic temporal deployment in the unity of the presentation.

Pictorial representation, therefore, shows that perception does not consist in the mere transformation of one thing into a structurally different something else (as happens in the standard epistemological stimulus/percept relation) or in a sum of two-dimensionally arranged pieces. Rather, pictorial representation is the expression of an intrinsically three-dimensional dynamic process (non-independent parts in 3D expansion in the TP), which it represents in a particular medium (the canvas) using certain structural rules of 'composition' shared by the two relative spaces.

Further deformations actuated in pictorial space display other components of natural perception, some of which have already been discussed:

1. The continuity of perceived visual space, also amodally.
2. The existence of a proprio-centric peripersonal space.
3. The specific concepts of perceptive direction, velocity and boundary.
4. The concept of 'pattern' in the sense used here.
5. The concept of surface in perceptive space.
6. The concept of tertiary qualities.

Boccioni's *The Street Enters the House*, for example shows simultaneity of points of view in actual perception, but with a further distinctive feature. The picture absolutizes the essential *proprio-centric point of view* in perception, because the woman on the balcony, who assumes the perceiver's point of view, is the attractor of the visual space, which is deformed in her direction and displays all the characteristics of tactility at decreasing distances. The maintenance of spatial and temporal unity is due to the (potential) gaze in different directions permitted by the position of the woman and by her profile and performs a dynamic and anticipatory function for other directions of the gaze. In this case, the completed form carries with it traces of the past and anticipation of the future deployment.

A similar deformation also constitutes the structural skeleton represented by Duchamp's *Nude Descending a Staircase*, where what is actualized, in the *sequence of phases* making up the TP, is the primitive of the *direction of becoming*, rightwards

and downwards. The painting is a strong figurative argument for the *presence of a simultaneity of phases* in seeing.

Other paintings that magnify this essential feature of seeing are, for example, Balla's *Dog on a Leash* or *Young Girl Running on a Balcony*, which show the forms of *temporal continuity* in visual space.

Balla also used deformation to represent the specificity of the characteristics of *direction, velocity and boundary of surfaces* which limit the volumes of visual appearances. What rules did Balla follow when painting *Abstract Velocity and Noise*? Analysis of the phenomenal depicted space reveals a number of interesting features. The painting intermodally represents the pattern 'velocity' and the pattern 'noise' in the same sensory modality, the visual one, exploiting the synaesthesia and similarity between two patterns with very similar syncopated rhythms. The idea of modelling this type of intermodal representation immediately prompts reflection on the current lack of a mathematical theory of perceptive continua able to account for the coincidence of boundaries, the multiformity and multiplicity of directions, and in short the intermodal structural skeleton of such depiction (Albertazzi 2002; Bell 2005).

Deformation in pictorial space is equally evident in the case of the depiction of surfaces and volumes in actual seeing. Consider Boccioni's *Cyclist*. Represented in this case is the *simultaneity of directions* in the perceived movement and the *dynamic between 3D and 2D parts* in the presentation of an event. What Boccioni puts into form is not 'the' cyclist, nor is it 'a' cyclist; rather, it is 'cycling' apprehended simultaneously in its unfolding in the TP.

A depiction more closely centred on the pattern as such – whether it pertains to the tectonic the global quality of the whole, or its physiognomic expressiveness – is Kandinsky's *Dominant Curve*. In this painting the objects depicted are the already-mentioned 'curvedness', 'acuteness', 'roundness', 'angleness', and so on, which assume the role of protagonists. One of these objects – 'curvedness' – dominates the others.

All the examples cited possess a 'demand character' tied to the specific patterns depicted. Artistic figuration, moreover, provides examples in which the expressive and emotional qualities themselves are the dominant pattern. This is the case of Dudreville, for instance, but it is even more strikingly exemplified by Giacometti's sculpture, *Femme Egorgée*.

In this case, once again, the structural skeleton is represented by *indexes* of it made manifest in the specific style adopted and through the medium selected. The drama of the event is conveyed by the distress caused by the perceived form, which evokes the *hidden visual form*. In other words, the *visual form is seen from the point of view of its structural skeleton*.

What, therefore, are the similarities and differences between the appearances of natural perception and the appearances of pictorial perception?

As to similarities, both natural and pictorial appearances are products of acts of intentional reference, in that:

1. They have a specific intrinsic direction towards the putting-into-form of an objectuality of some type.
2. The continuity of seeing constructs and connects them together as acts of different quality (for example acts of presentation with acts of representation in pictorial space, i.e. acts founded on previous ones).
3. They are endowed with both a visible and hidden structure.

As to differences, the already-mentioned distinction between *perceptive presence* and *mental presence* applies, or the fact that:

1. Seeing in natural perception is the *perceptive presentation* of a present object (i.e. non-independent, coexistent, and therefore evident).
2. Seeing in a pictorial depiction is the *mental presentation* of an absent object (i.e. the original object of intentional reference) represented through the content expressed (see Introduction, §5).
3. The two acts of vision involve *different* forms of subjective completion.

In the case of seeing *in* pictures, therefore, a perfect unfolding of the event does not occur; rather the event is deformed twice (i) by the representation and (ii) by the mode of its location in the space represented, that is, by the specific style of the work (see Ingarden 1962, Chapter 6; Husserl 1980).

Abstract paintings, I believe, best exemplify the structure of phenomenal appearances as they arise in the microgenesis of TP: the non-independent components of the whole in question, such as its pattern (colour and shape), direction, position, tension, etc., according to the terminology used here.

Towards a conclusion

I have analysed in particular the *structure of visual appearances* (as ‘phenomenal’ phenomena), claiming that they pertain to the mind (i.e. psyche) and not to the brain (the opposite view is taken by Zeki 1999). I have underlined that they are *spatio-temporal unfolding structures* with their own primitives. As such, they pertain to an emergent level of reality situated at the boundary of inner psychophysics, which is the real border of consciousness.

As noted in the introduction, the structure of visual space has a complexity that cannot be reduced to physical, psychophysical or neural features, *despite the presence and reciprocal dependence of these and other levels in the organization of the visual field*. Visual space, in fact, is a highly subjective space completed by cognitive operations of various kinds.

As to the problem of *continuity* or *discontinuity* between seeing and thinking, between modal and amodal perception – a problem left unresolved by Kanizsa's analysis – the difference between the space of natural perception and pictorial space furnishes some useful information.

Perceptive seeing is constructed in seeing the object through its *visual parts* (moments). Seeing in pictorial depiction is constructed in seeing the object *in* and *through* the *visual parts of the picture* – that is, of the *content represented*. Act and parts of the act are different in the two cases. In the complex architecture of the relations among act, object and content, in fact, the term 'represented' has two different meanings: on the one hand it denotes the parts of what is represented in the sense of 'content'; on the other, it denotes what is represented in the sense of 'object' (Twardowski 1977). Pictorial space is therefore the result of a *double form of intentional reference*: for example, the proprio-centricity of perceived space becomes hierarchically multiple – as shown by Hildebrand when he analysed the symmetry of the frontal surface (Hildebrand 1961). Once again, moreover, one should not confuse the *form brought to completion* in pictorial space with the *putting-into-form of painting*: from the latter point of view, every painting is a dynamic visual experience (Ione 1999).

Many art works of the twentieth-century, precisely because they *put-into-form* the various *components of the structural skeleton* of an object of intentional reference – aid understanding of the complex nature of phenomenal appearances as a case of *reverse intentionality* (Ingarden 1962, Chapter 10). The structural skeleton, in fact, is not *inferred* from the actual form – as demonstrated by Arnheim's example of a triangle transformed into other, different triangles by moving an angle point (Arnheim 1954:91). The structural skeleton is *visible* in the forces and the *inner relations* that regulate the visual appearance.

We are still only at the beginnings of a theory of appearances able to account for the laws of dependence among the various levels of reality and also able to model them. For example, such a theory would describe the geometry of cognitive space and of its unfolding in diverse spaces embedded in each other: a Herbartian geometry based on the idea of *geometric imagination* as the locus for visualizing and/or presenting *serial forms*.

From what I have said, therefore, visual and pictorial space are *two layers of the same ontological level* of appearances: *both* of them are the places of intentional objects. In both, in fact, the continuum takes ontological precedence over discrete structures. From this point of view, visual space and pictorial space are the 'intentional canvases' of two kinds of unfolding 'drawings'. Depero wrote: 'As I think, so I paint'.

Notes

* From a report by a student at the Bauhaus, cit. in Whitford (1984:99).

1. Consider Klee's painting '*Blaue Nacht*', which depicts in pictorial terms certain aspects of Wertheimer's laws of 1923, in particular the continuity of good form, the factor of direction, and degrees of *Prägnanz*; or his lectures collected in his Notebooks on overlapping planes cited by Wertheimer and Fuchs. See Klee (1961: 117); Teuber M. (1976); Teuber H.-L. (1961). See also Neu (1978).
2. Note that the original title of Klee's *The Thinking Eye* was *Das bildnerische Denken* (1956), which conveys better the idea of the pictorial nature of thought.
3. The term derives from *phainesthai* (*to appear*) and it is obviously connected to the phenomenal (not phenomenalist!) concept of appearances (*phantasmata*), principally in the visual field. But by phantasia is generally meant a capacity to visualize and/or to present. In modern terms, it concerns the question of cognitive integrations at primary level (see Albertazzi 2003).
4. This concept at the basis of James's 'flow of consciousness' and Husserl's double intentionality. See James (1890); Husserl (1966a).
5. The term 'descriptive' is taken from Brentano (1995) and denotes analysis of the act, content and object of the psychic phenomenon.
6. By 'inner relation' is meant, not a Russellian type of relation (aRb) in which individuals maintain their integrity but a type of relation in which the components become non-independent parts of the whole. An example is the relation among the parts of a melody, which can be entirely replaced while preserving the tonal distances among the notes (principle of the invariance of gestalt configurations).
7. The concept of *Prägnanz* has two variants: that of a 'characteristic that makes a form good, symmetric and simple'; and that of 'point in a series which makes it into an excellent (*ausgezeichnet*) singularity'. Criticisms of the original concept of *Prägnanz* were made internally to *Gestaltpsychologie* by Metzger and Rausch, until it became a tendency by the field towards equilibrium, also at the expense of individual *Gestalten* with little pregnancy. See Kanizsa, Luccio (1985).
8. In the theory of intentional reference the act of reference performs three functions: it notifies a mental content (*Kundgabe*), triggers a similar content (*Auslösung*), and represents it (*Darstellung*) with expressive means. See Brentano (1956).

References

- Albertazzi, L. (1997). Continua, adjectives and tertiary qualities. *Axiomathes*, 8, 7–30.
- Albertazzi, L. (1998). The aesthetics of particulars. A case of intuitive mechanics. *Axiomathes*, 9, 169–196.
- Albertazzi, L. (2002). Towards a neo-Aristotelian theory of continua. In L. Albertazzi (Ed.), *Unfolding Perceptual Continua* (pp. 29–79). Amsterdam & Philadelphia: Benjamins.
- Albertazzi, L. (2003). From Kanizsa back to Benussi. Varieties of intentional existence. In L. Albertazzi (Ed.), *The Legacy of Kanizsa in Cognitive Science* (pp. 239–259). Special issue of *Axiomathes* 13.
- Albertazzi, L. (2004). Stereokinetic shapes and their shadows. *Perception*, 33, 1437–1452.

- Albertazzi, L. (2005). *Immanent Realism. Introduction to Brentano*. Berlin-New York: Springer.
- Albertazzi, L., Jacquette, D., & Poli, R. (Eds.). (2001). *The School of Alexius Meinong*. Ashgate: Aldershot.
- Arnheim, R. (1954). *Art and Visual Perception. A Psychology of the Creative Eye*. Berkeley: The Regents of the University of California.
- Arhneim, R. (1969). *Visual Thinking*. Berkeley: The Regents of the University of California.
- Bonaventura, E. (1929). *Il problema psicologico del tempo*. Milano: Soc. An. Istituto Editoriale Scientifico.
- Bell, J. (2005). *The Continuous and the Infinitesimal in Mathematics and Philosophy*. Milan: Polimetrica.
- Benussi, V. (1906). Über Vorstellungsinadequätheit. I. Das Erfassen gestaltnmehrdeuticker Komplexe. *Zeitschrift für Psychologie*, 42, 22–25.
- Benussi, V. (1913). *Psychologie der Zeitauffassung*. Heidelberg: Winter.
- Brentano, F. (1956). *Die Lehre vom richtigen Urteil*, ed. by F. Mayer-Hillebrand. Bern: Francke.
- Brentano, F. (1995). *Descriptive Psychology*, ed. by B. Müller. London: Routledge.
- Brentano, F. (1988). *Philosophical Investigations on Space, Time and the Continuum*. London: Croom Helm.
- Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature neuroscience*, 7 (3), 308–313.
- De Vries, L. (1971). *Victorian Inventions*. London: John Murray.
- Ehrenfels, Ch. von (1890). Über Gestaltqualitäten. *Vierteljahrschrift für wissenschaftliche Philosophie*, 14, 249–292.
- Fraisse, P. (1964). *The Psychology of Time*. London: Eyre and Spottiswood.
- Fraisse, P. (1974). *Psychologie du rythme*. Paris: PUF.
- James, W. (1890). *Principles of Psychology*. Boston: Holt and Co.; repr. New York: Dover Publications 1950.
- Hecht, H., Schwartz, R., & Atherton, M. (Eds.). (2003). *Looking Into Pictures*. Cambridge, MA: MIT Press.
- Herbart, Johannes F. (1816). *Lehrbuch der Psychologie*. Reprint 1965. Amsterdam: E.J.Bonset.
- Hildebrand, A. von (1861). *The Problem of Form in Painting and Sculpture*. New York: Stechert.
- Husserl, E. (1913). *Ideen zu einer reinen Phänomenologie und phänomenologische Philosophie*. Halle: Niemeyer. Repr. 1950, ed. by W. Biemel (I), M. Biemel (II), & M. Biemel (III), Husserliana III. Den Haag: Nijhoff, Husserliana IV, ivi, 1952, Husserliana V, 1952.
- Husserl, E. (1900/1). *Logische Untersuchungen*. Halle: Niemeyer. Rep. 1975, vol. I ed. by E. Holenstein, Husserliana XVII. Den Haag: Nijhoff. Rep. vols. I & II 1984, ed. by U. Panzer, Husserliana XIX, ivi, 1, 2.
- Husserl, E. (1966a). *Zur Phänomenologie des inneren Zeitbewusstseins*, ed. by R. Boehm, Husserliana X. Den Haag: Nijhoff.
- Husserl, E. (1966b). *Analysen zur passiven Synthesis*. The Hague: Nijhoff.
- Husserl, E. (1980). *Phantasie, Bildbewusstsein, Erinnerung: zur Phänomenologie der anschaulichen Vergegenwärtigungen: Texte aus dem Nachlass (1898–1925)* ed. by E. Marbach. Den Haag: Nijhoff.
- Ingarden, R. (1962). *Untersuchungen zur Ontologie der Kunst: Musikwerk, Bild, Architecture, Film*. Tübingen: Niemeyer.
- Ione, A. (1999). An inquiry into Paul Cézanne. The role of the artist in studies of perception and Consciousness. www.imprint-academic.com/jcs.
- Itten, J. (2002). *Elemente der bildende Kunst. Studienausgabe des Tagebuchs* (2nd ed.). Leipzig: Seemann Verlag.

- Kandinski, W. (1926). *Punkt Linie zur Fläche*. Bern: Benteli.
- Kanizsa, G. (1991). *Vedere e pensare*. Bologna: Il Mulino.
- Kanizsa, G., & Luccio, R. (1985). La pregnanza e le sue ambiguità. *Psicologia italiana*, 7, 11–39.
- Katz, D. (1906). Experimentelle Beiträge zur Psychologie des Vergleichs im Gebiete des Zeitsinns. *Zeitschrift für Psychologie*, 42, 302–340.
- Koffka, K. (1935). *Principles of Gestalt Psychology*. New York: Harcourt.
- Koenderink, J. J. (1990). *Solid Shape*. Cambridge, MA: MIT Press.
- Koenderink, J. J., & van Doorn, A. (2003). Pictorial space. In H. Hecht et al. (Eds.), *Looking into Pictures*. Cambridge, MA: MIT press.
- Kenkel, F. (1913). Untersuchungen über den Zusammenhang zwischen Erscheinungsgrösse und Erscheinungsbewegung. *Zeitschrift für Psychologie*, 67, 358–449.
- Klee, P. (1924). On Modern Art. In P. Klee (Ed.), *The Thinking Eye*. London: Lund Humphries.
- Klee, P. (1961). *The Thinking Eye*. London: Lund Humphries.
- Kopferman, H. (1930). Psychologische Untersuchungen über die Wirkung zweidimensionaler Darstellungen körperlicher Gebilde. *Psychologische Forschung* 13, 293–364.
- Leyton, M. (1992). *Symmetry, Causality, Mind*. Boston: MIT Press.
- Libet, B. (1982). Brain stimulation in the study of neuronal functions for conscious sensory experience. *Human Neurobiology*, 1, 235–242.
- Libet B., Wright, E. W., Feinstein, B., & Pearl, D. K. (1979). Subjective referral of the timing for aconscious sensory experience. *Brain*, 102, 191–224.
- Lipps, Th. (1897). *Raumaesthetik und geometrisch-optische Täuschungen*. Leipzig: Barth.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman.
- Massironi, M. (2002). *The Psychology of Graphic Images*. London: Elsevier, Erlbaum.
- Meinong, A. (1899). Über Gegenstände höherer Ordnung und deren Verhältniss zu inneren Wahrnehmung. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 21, 182–272.
- Meinong, A. (1904). *Untersuchungen zur Gegenstandstheorie und Psychologie*, ed. by A. Meinong. Leipzig: Barth.
- Metzger, W. (1941). *Psychologie*. Darmstadt: Steinkopff.
- Michon, J. (1967). *Timing in Temporal Tracking*. Assen: Van Gorcum.
- Michon, J. A., & Jackson, J. (1985). *Time, Mind and Behaviour*. Berlin-New York: Springer.
- Michotte, A. et al. (1962). *Causalité, permanence et réalité phénoménale*. Louvain: Publications Universitaires.
- Musatti, C. L. (1924). Sui fenomeni stereocinetici. *Archivio Italiano di Psicologia*, 3, 105–120.
- Musatti C. L. (1955). La stereocinesi e la struttura dello spazio visibile. *Rivista di Psicologia*, 49, 3–57.
- Neu, T. (1978). *Von der Gestaltungslehre zu den Grundlagen der Gestalt*. Ravensburg: Otto Mayer Verlag.
- Rensink, R. A. (2000). Seeing, sensing, scrutinizing. *Vision Research*, 40, 1469–1487.
- Rosen, R. (1985). *Anticipatory Systems. Philosophical, Mathematical and Methodological Foundations*. New York: Pergamon Press.
- Rosenthal, V. (2004). Microgenesis. Immediate experience on visual processes in reading. In A. Carsetti (Ed.), *Seeing, Thinking & Knowing* (pp. 221–243). Dordrecht: Kluwer.
- Sander, F. (1930). Structures, totality of experience and gestalt. In C. Murchison (Ed.), *Psychologies of 1930* (pp. 188–204). Worchester, MA: Clark Un. Press.
- Stern, W. (1897). Über psychische Präsenzzeit. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 13, 325–349.
- Stern, W. (1898). *Psychologie der Veränderungsauffassung*. Breslau: Preuss und Junger.

- Thinès, G. (1977). *Phenomenology and the Sciences of Behaviour. An Historical and Epistemological Approach*. London: Allen & Unwin.
- Teuber, H.-L. (1961). Perception. In H. W. Magoun (Ed.), *Handbuch of Physiology, Section 1, Neurophysiology*, 3. Washington, DC: American Psychological Society.
- Teuber, M. (1977). Blue Night by Paul Klee. In M. Henle (Ed.), *Vision and Artifact* (pp. 131–151). New York: Springer.
- Twardowski, K. (1977). *On Content and Object of Presentations. A Psychological Investigation*. The Hague: Nijhoff.
- Ulanowicz, R. E. (1997). *Ecology. The Ascendent Perspective*. New York: Columbia University Press.
- Vishwanath, D. (2005). The epistemological status of computational vision and vision science and its implications for design. *Axiomathes*, 15 (3), 399–480.
- Werner, H., & Wapner, S. (1952). Toward a general theory of perception. *Psychological Review*, 59, 324–338.
- Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. *Psychologische Forschung*, 4, 301–350.
- Wirth, W. (1934). Zu den Bewusstseinsphänomenen der Gestaltauffassung. *Archiv für die gesamte Psychologie*, 91, 507–538.
- Withford, F. (1984). *Bauhaus*. London: Thames & Hudson.
- Zanforlin, M. (2000). The various appearances of a rotating ellipse and the minimum principle: A review and an experimental test with non-ambiguous percepts. *Gestalt Theory*, 22, 157–184.
- Zeki, S. (1999). *Inner Vision. An Exploration of Art and the Brain*. Oxford: Oxford University Press.

Rudolf Arnheim's graphic equivalents in children's drawings and drawings and paintings by Paul Klee

John Willats

Until recently theories of depiction were based on 'camera' theories of visual perception and 'copying' theories of picture production. Together these led to the belief that pictures are more or less direct copies of mental images that take the form of two-dimensional views. One problem with this account is that it cannot explain the apparent anomalies in drawings by young children, such as the lack of perspective and the inclusion of transparencies. Another problem is that copying theories cannot account for somewhat similar anomalies in the work of twentieth-century avant-garde artists such as Braque, Kandinsky, Klee, Matisse and Picasso. Rudolf Arnheim (1956, 1974) rejected the copying theory of depiction and proposed instead that pictures provide graphic equivalents within a given medium for the objects they depict. One of the strengths of Arnheim's approach was that he applied it to analyse the representational systems in both children's drawings and artists' pictures – topics that are almost always treated in isolation by other writers.

Arnheim's work has been both fruitful and influential. However, his approach was limited by being based on Gestalt psychology, which lacked the technical vocabulary to describe pictorial structures in any detail. As a theory of visual perception Gestalt psychology has now been replaced by more modern theories, notably that developed by David Marr (1982). Nevertheless, Arnheim's idea of graphic equivalents still has great value, and as a result of new work in vision science and artificial intelligence it is now possible to describe these graphic equivalents in terms of three related representational systems: the *drawing systems*, the *denotation systems* and the *mark systems*.

What I have called the copying theory of pictures has had a long history, and in the classical Greek period took the form of the theory of *mimesis*. In this period a distinction was made between seeing and knowing similar to that made by the early writers on children's drawings:

In the classical period and commonly in the ancient tradition *aisthesis* was described as the processes in which mental images of contingent qualities and shapes of individual things are presented to the mind. When a person sees a house, for example, there is a mental image of the house in the mind of the person, an image of its colours and shapes. The basic metaphor used to characterize this process was that of pressure. An individual thing presses its contingent qualities and shapes upon the senses like a signet ring which, when stamped into wax, delivers its form but not its matter to the wax. However, when we think 'house' the mind entertains the essence, real nature or the 'houseness of houses' which is something general and not accessible to the senses (*aisthesis*) but only to thought (*noesis*) since thoughts do not have individual and contingent properties. (Sörbom 2002: 20)

With the advent of optical theories of perception during the Renaissance the metaphor of pressure was replaced by the metaphor of the *camera obscura*. In a *camera obscura* the light coming from the scene forms an image on a screen in much the same way, it was thought, that images are formed on the retina of the eye. Artists had used *camera obscura* images as an aid to producing pictures in perspective since at least the sixteenth century, and this was used as a model for picture production. In a *camera obscura* the light coming from the scene is projected on to a screen, and the artist traces round this image to produce a drawing which is necessarily in perspective (Kemp 1992). It thus seemed natural to suppose that artists could produce pictures in much the same way, by copying perspectival internal images.

J. J. Gibson (1978) described this traditional theory (which he himself rejected) in the following terms: 'Drawing is always copying. The copying of a perceptual image is drawing from life. The copying of a stored image is copying from memory. The copying of an image constructed from other memory images is drawing from imagination.' (p. 230). As Gibson pointed out, there are a number of difficulties with this theory, not the least of which is that humans have no means of projecting internal images on to the picture surface in the way that *camera obscuras* project images on to a screen. But as Gibson said, 'If you cannot trace around the projected mental image at least you can copy it freehand. Perhaps drawing is not exactly like this, they say, but *something* like this. Otherwise what could it be?' (Gibson 1978: 230–231).

The copying theory of pictures is very compelling, and appears to work well in explaining realistic paintings and drawings; but the snag is that it cannot be used to explain either children's drawings or the drawings and paintings of the twentieth-century avant-garde. Even when children are asked to draw scenes or objects from life the geometry of the drawings they produce rarely resembles that of their view of the scene.

Ever since the end of the nineteenth century the most popular way of explaining drawings of this kind has been to say that 'young children draw what they know

and older children draw what they see'. Clark (1897), for example, asked children aged 6 to 16 to draw an apple with a pin pushed through it, and found that the younger children included the whole of the pin in their drawings even though part of it was out of sight. Clark asked, 'Why do not the six-year-olds draw things as they appear by the unconscious power of imitation, for which they are so famous?' His answer was, 'The facts are they draw things as they are known to be, not as they appear' (p. 287). This explanation, or variants of it, persisted all through the twentieth century and still forms the basis of most generally accepted theories of children's drawings. There are a number of problems with this account (Willats, *in press*), but the most serious of these is that a simple two-stage theory cannot account for the richness and complexity of the changes in children's drawings that take place with increasing age.

Almost alone among writers on children's drawings Arnheim rejected the copying theory of drawing and the distinction between knowing and seeing.¹ He considered that the principle determining developmental change was that of progressive differentiation, so that all forms of development tended to proceed from simple to complex. Applying this general law to drawing development, he first proposed that: 'Any shape will remain as undifferentiated as the draftsman's conception of his global object permits'. For example, he considered that the head/body of a tadpole figure such as that shown in Figure 3 typically consists of a more or less circular form because this captures the child's earliest concept of the human figure as a single extended volume or lump. Secondly, he proposed that 'Until a visual feature becomes differentiated, the total range of its possibilities will be represented by the structurally simplest among them' (Arnheim 1974: 181). Thus, even though children know that the head/body of a human figure is elongated it will be represented by a more or less circular form, because this is the simplest version of a range of closed curved forms which would include not only circles but ellipses of varying proportions. Similarly, in their earliest drawings, children use the same more or less circular form to represent houses and cubes. This is not because they do not know that houses and cubes have flat faces, but rather that a circular form is the structurally simplest way of representing a volume which is about equally extended in all directions. What I shall have to say about children's drawings is very much in the spirit of these principles, but I shall show how they can be applied in more detail, using a more precise vocabulary than that which was available to Arnheim.

Much of this vocabulary is derived from the work of David Marr. According to Marr and Nishihara (1978) all systems for representing shape and space must have at least two components, and this applies to both internal descriptions of objects and scenes and representations of such objects and scenes in the form of pictures. The first of these components is a co-ordinate system: that is, a way of describing the spatial relations between features. These co-ordinate systems need

a frame of reference, and this may be either what Marr called object-centered or viewer-centered. In viewer-centered co-ordinate systems spatial relations are given relative to the viewer's line of sight, but in object-centered descriptions spatial relations are given relative to the principal axes of the object or scene, independently of the position from which they are viewed.

The other component that all representational systems must have is the units or features of which they are composed. These primitives may be zero-, one-, two- or three-dimensional. Retinal images, for example, are necessarily built up of zero-dimensional primitives because the receptors can only take in information about the light coming from the scene in the form of point intensities. However, Marr argued that it would be very uneconomic to store descriptions of objects in this form. Instead he proposed that the primary function of the human visual system was to take the ever-changing viewer-centered descriptions available at the retina and use them to build more stable object-centered descriptions that were independent of lighting conditions or any particular point of view. These descriptions are then stored in long term memory and used to recognise objects when we see them again under new lighting conditions or from a new direction of view. Marr called such descriptions '3D models', and argued that they are built up of three-dimensional or volumetric primitives. An internal 3D model of a table, for example, might take the form of a flat volume or 'slab' standing on four long volumes or 'sticks', rather than being composed of one-dimensional edges or two-dimensional surfaces.

In addition, early work in artificial intelligence on the automatic analysis of line drawings by computers (Clowes 1971; Huffman 1971), had shown how important it was to use a different vocabulary to describe the features of scenes from that used to describe the corresponding features in pictures. In everyday speech we often speak of pictures as if they contained objects such as tables and people, or edges and corners, but in reality pictures are made up of picture primitives such as lines and line junctions that stand for or *denote* such features.

This early work in artificial intelligence, together with the idea of the two components of representational systems – the primitives of which they are composed and the spatial relations between them – led to the concept of the drawing systems and the denotation systems. The *drawing systems* are systems such as perspective and oblique projection that map spatial relations in the scene into corresponding spatial relations in the picture. In drawings in perspective, for example, front-to-back directions in the scene, such as the parallel directions of the side edges in an object-centered description of a table, are represented by converging directions on the picture surface.

In oblique projection, in contrast, the side edges are represented by parallel oblique lines.

The *denotation systems* are systems such as line drawing that map scene primitives into corresponding picture primitives. In a line drawing of a table, for

example, the edges of the table are represented by lines, and the corners by various types of line junctions. There are two other types of common denotation systems: optical systems, and systems based on regions as picture primitives. In optical systems such as photographs and Impressionist and Pointillist paintings zero-dimensional primitives or points are used to represent the hues and intensities of the cross-sections of light rays coming from the scene. In region-based systems two-dimensional primitives or regions are used to represent either the three-dimensional volumes of objects, or their two-dimensional faces. Examples of pictures in region-based systems include early Greek vase paintings, works by Picasso and Matisse, and children's early drawings.

A full account of the representational systems in pictures must also include an account of the *mark systems* on which they are based. Picture primitives are abstract entities, and in real pictures must be represented by physical marks. In a pen and ink drawing of a person or a table the picture primitives would be represented by traces of ink on paper, but in other versions of such pictures they might be represented by lines of mosaic or stitches in a tapestry.

The main reason why young children's drawings often look so strange is that young children use representational systems that are radically different from those used by adults. If we are to make any sense of children's drawings the first questions we have to ask are: 'What are the drawing, denotation and mark systems on which they are based? Such questions can rarely be answered by inspecting single drawings in isolation. Often the best way of answering these questions is to look at the way in which children's representational systems change as they get older, and especially at the anomalies in their drawings that arise at certain stages of development (Willats, in press).

Even after the lapse of nearly a century the drawings and paintings of the avant-garde artists of the early twentieth century can also look strange, and I shall suggest that this is because they are often based on representational systems that are different from those used by the academic painters of the nineteenth century. In particular I shall show that many of Paul Klee's drawings and paintings are based on systems that are more like those used by young children than the systems used by the academic painters who preceded him. However, I shall also argue that, unlike children, Klee used these systems quite deliberately as a way of investigating the nature of depiction.

It used to be thought that young children's scribbles simply consisted of marks made at random, or that at best they were rehearsing the motor movements they would later need when drawing proper began. It now seems likely, however, that even these early scribbles can contain quite complex representational structures (Matthews 1999). Figure 1 shows a drawing of a man and Figure 2 a drawing of a house, both produced by a girl aged 2 years 6 months. Almost the only spatial relation represented in these early drawings is the topological relation of *spatial*

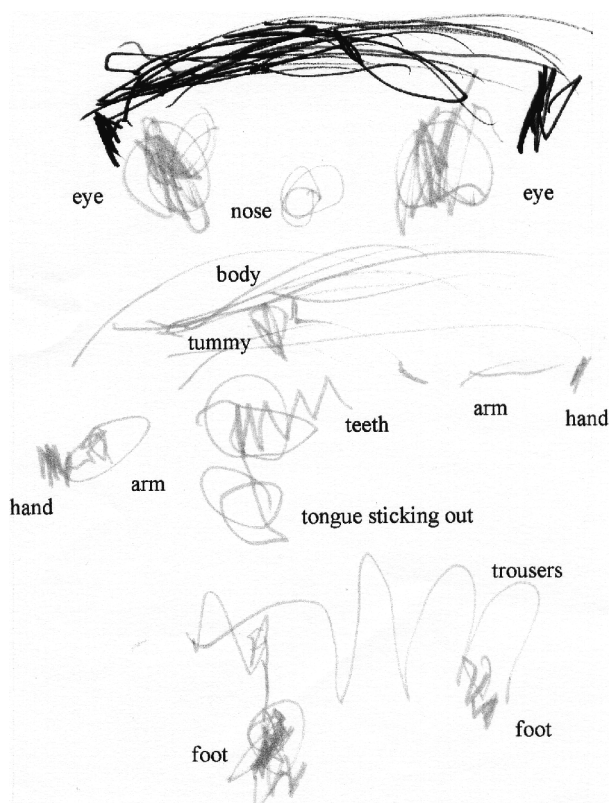


Figure 1. Megan Erridge, 2 years 6 months. A drawing of a man. The spatial order of the features is roughly correct but they are not connected. The hair is represented by a long patch of scribble and the eyes, nose, tongue and feet are represented by round patches. Collection of the author.

order. In Figure 1 the man's hair is shown at the top of the page, the feet at the bottom, the eyes on either side of the nose, and the hands and feet on either side of the vertical central axis. Similarly, in the drawing of a house shown in Figure 2 the roof and chimney are shown at the top and the door at the bottom, with the windows on either side. One other topological relation also appears in this drawing, however, that of *enclosure*: the door handle is shown within the region representing the door.

The only shape property represented by the picture primitives is that of *extendedness*. Round features in the scene such as the eyes in Figure 1 and the windows in Figure 2 are represented by round regions in the picture, and long features in the scene such as the man's hair and the roof of the house are represented by lines or long regions. In this context it is important to realise that 'round' does not necessarily mean 'circular'. Extendedness is the most basic of all shape properties, and

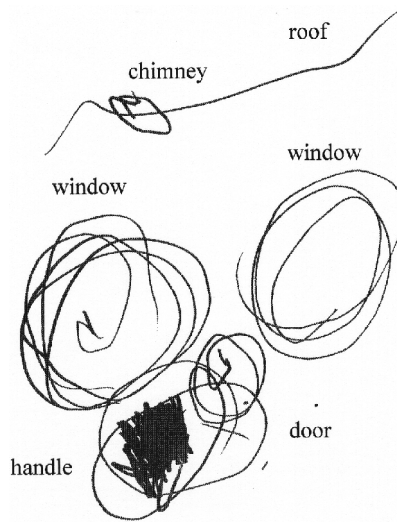


Figure 2. Megan Erridge, 2 years 6 months. A drawing of a house. The spatial order of the parts is roughly correct, and the handle is enclosed by the door. The roof is represented by a long line and the chimney, the windows, the door and the handle are represented by round patches. Collection of the author.

simply refers to the saliency of extension of the scene primitives in one, two or three directions in space and the saliency of the corresponding picture primitives in one or two directions on the picture surface (Willats 1997).

Finally, the mark system in these drawings is one in which patches of scribble, made up of lines, are used to define the regions. The lines in these scribbles may be more or less densely spaced, and single lines are occasionally used to represent individual features, but the marks in these early drawings are predominantly patches of scribble.

In the next stage of development this mark system undergoes a dramatic change, exemplified by the tadpole figure shown in Figure 3 and the drawing of a house shown in Figure 4. The picture primitives are still regions, but these regions are defined by outlines enclosing areas rather than patches of scribble. There are several advantages in making this change, the main one being that the outlines of the regions can be interpreted as the contours of smooth forms, making the objects depicted much more immediately recognisable. Another advantage is that patches of scribble can now be used to represent textured features, such as the smoke coming out of the chimney of the house.

In terms of picture production, however, the denotation system in these drawings remains much the same as it was in the drawings shown in Figures 1 and 2. Round regions are used to denote round volumes such as the head or head/body of the tadpole figure and the house and single lines (representing notional long re-

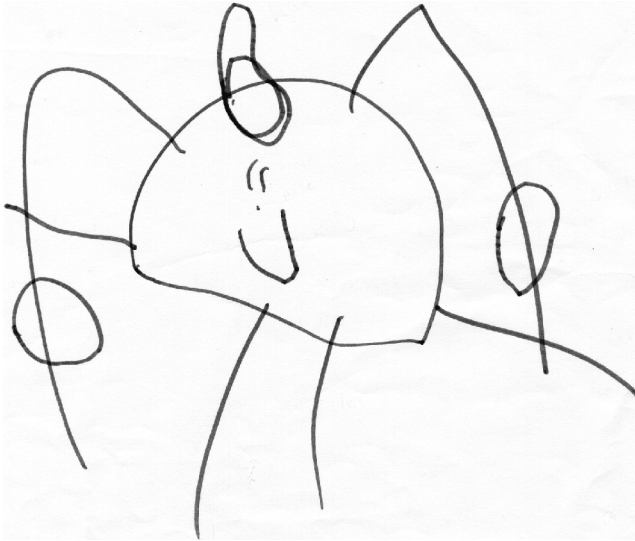


Figure 3. Adele Palmer, 3 years 6 months. A Little Girl in a Hat. The head or head/body, the hat and the bunches of hair are represented by round regions, which in turn are represented by areas enclosed by an outline. Collection of the author.

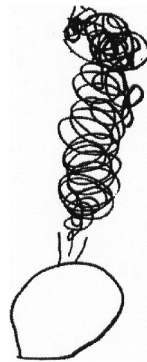


Figure 4. Ben Matthews, 3 years 3 months. A drawing of a house (detail of Father Christmas on His Sleigh with Reindeer). The house is represented by a round region, which in turn is represented by an area enclosed by an outline. The smoke is represented by a patch of scribble. Taken from Matthews (2003, Figure 53). Courtesy of J. Matthews.

gions) are used to denote long volumes such as the arms and legs of the figure. The drawing system in these drawing still includes the topological relations of spatial order and enclosure, but to these has been added the ‘join’ relation. That is, lines or regions that are joined in the picture represent features that are joined in the scene.

As a result of these changes the drawings shown in Figures 3 and 4 are far less ambiguous than those shown in Figures 1 and 2, and individual features, which could only be recognised by adding their names in the earlier drawings, can now be recognised without much difficulty. As a result these drawings are far more effective as representations: that is, the objects they are intended to represent can be seen, however crudely, in the picture.

The next developmental change that takes place is that the vocabulary of picture primitives, previously confined to long regions, round regions and dots, is enriched by varying the proportions of the regions and by the addition of shape modifiers. These *shape modifiers* are used to represent nonaccidental shape features in the scene such as 'being bent' or 'having flat faces' (Hollerbach 1975). In Figure 5 finer distinctions have been made among the regions in respect of their extendedness: the head is represented by a circular region and the arms and legs, previously represented by marks in the form of single lines, are now represented by long regions. In addition the legs and feet are not represented by separate regions, as they would be in earlier drawings; instead, they are represented together by single elongated regions that are bent at the ends. A region with flat sides is used to represent the cylindrical shape of the hat, and the hat brim, a flat volume or disc, is represented by a line. The brim is shown touching the head rather than partially occluding it.

In Figure 6 the body of the house is represented by a square region. This is, in effect, a round region used to represent the whole volume of the house, with the addition of the shape modifier 'having flat sides' used to represent the shape property of 'having flat faces'. A common feature of drawings of this kind is that the windows are shown attached to the bounding outline. If the square region represents the whole volume of the house, then this outline can be considered to represent the exterior surface, so the windows are shown where they belong, attached to this surface. In the scene the house is surrounded by snow, and is therefore represented surrounded by snow in the picture – a topological relation rather than a view.

Apart from the use of long areas rather than single lines to represent long regions, the mark system in these drawings remains essentially the same as it was in the drawings shown in Figures 3 and 4. Although most of the marks are lines, these lines do not represent edges such as the contours of the figure or the edges of the front of the house, but only serve to define the shapes of the regions.

In the next stage of development children use different denotation systems for representing smooth objects from the ones they use for representing objects with flat faces. In the drawing of a man shown in Figure 7 the picture primitives are still regions and these regions still denote whole volumes. Instead of these regions being shown separately, however, as they are in the drawings of a man shown in Figure 5, they are joined by continuous smooth outlines, a technique known as 'thread-



Figure 5. Boy, 6 years 8 months. A man with a hat. The arms and legs are subject to the shape modifier 'being bent'. Collection of the author.

ing'. These smooth outlines can be thought of as shape modifiers representing the shape property 'having a smooth surface'.

In contrast, the denotation system children now use for objects such as tables and houses changes to one in which the regions stand for individual faces rather than whole volumes. Moreover, the shapes of these regions correspond to the true shapes of the faces they represent, rather than to the shapes of the faces as they might appear from some particular point of view. To use Marr's terminology, they correspond to the shapes of faces in object-centered descriptions rather than viewer-centered descriptions. In Figure 8 the tabletop is shown as a rectangle

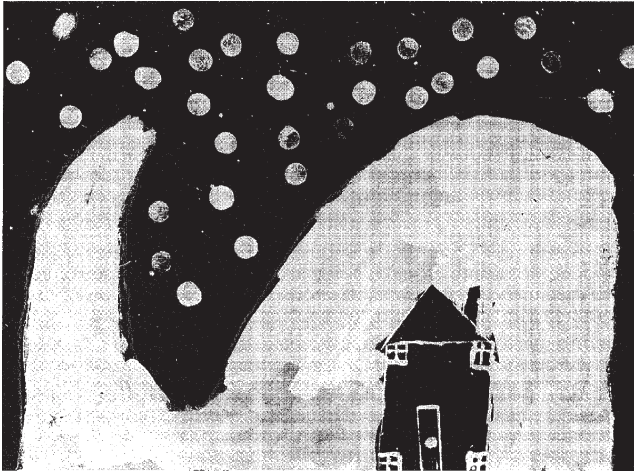


Figure 6. Arfan Kahn, 7 years 5 months. A House with a Huge Snowdrift. The house is represented by a round region subject to the shape modifier 'having straight sides'. The windows are shown touching the outline representing the surface of the house, and the house itself is surrounded by snow. Collection of the author.

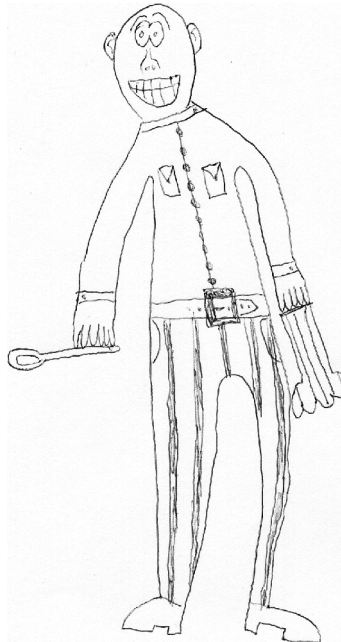


Figure 7. Boy, 11 years 3 months. A man holding a spoon and a chef's hat. The fingers are only touching the spoon and the hat, suggesting that the picture primitives are still regions. However, these regions are not represented separately but are joined by threading. Collection of the author.

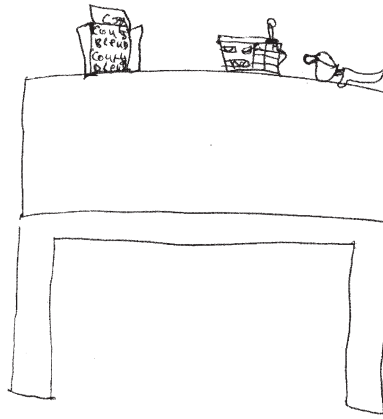


Figure 8. An example of a table with various objects on it drawn from a view of a real scene (Willats 1977a). This drawing is in vertical oblique projection, with regions as picture primitives representing faces as true shapes. The box is in the form of a fold-out drawing. The mean age of children producing drawings of this type was 11 years 11 months. Taken from Willats (1977b, Figure 5). Collection of the author.

(as it would appear in an object-centered description of a table) rather than as a trapezium as it would in a perspectival view.

The use of this denotation system has two consequences, however, both of which result in the characteristic anomalies that very often appear in the drawings that children produce during this developmental stage. The first is that it can be difficult or impossible to get the edges and corners of rectangular objects to join up properly. In the drawing of the box standing on the tabletop four faces of the box are shown as true shapes, so that it has proved impossible to get the corners to join up properly. Nicholls and Kennedy (1992) called these fold-out drawings, and Luquet (1927/2001) referred to this effect as *rabattement*. The second consequence of using this denotation system is that it makes children reluctant to show one object overlapping another because this would compromise the true shapes of the regions representing faces. As a result the objects on the table in Figure 8 have been shown side by side rather than overlapping each other, and in addition are shown perched on the far edge of the tabletop. In this way the regions representing the faces of the three objects and the tabletop can all be shown as true shapes.

Finally, children change to a new denotation system in which lines are used to denote edges and contours, and the marks in the form of lines in these drawings represent lines as picture primitives. Figure 9 shows a drawing of a man holding a hat. Although the parts of the body are still to some extent represented by separate regions, the boundaries of these regions are represented by lines that are beginning to represent genuine contours. Moreover, this drawing contains two types of



Figure 9. Girl, 10 years. A man holding a hat. The picture primitives are lines representing contours. The fingers are partially occluding the brim of the hat and T-junctions are used to represent the points of occlusion. One trouser leg partially occludes the other and its contour ends in an end-junction. Collection of the author.

line junctions that are characteristic of true line drawings. In Figure 7 the man's fingers are merely touching the hat and the spoon, which suggests that all these features have been conceived as separate regions. In Figure 9, in contrast, the hand is shown partially overlapping the hat, and there are two T-junctions representing points of occlusion where the edge of the hat brim disappears behind the fingers. In addition, one trouser leg is partially occluded by another, and the contour of the nearer leg correctly ends in an end-junction.

Figure 10 shows a representation of a table in the drawing system known as oblique projection. Oblique lines are used to represent the side edges of the table, so that for the first time directions in the third dimension of the scene can be unambiguously represented in the picture. Using lines rather than regions as picture primitives also enables the shapes of the regions representing features such as the tabletop to be distorted out of their true shapes, so that the problems arising in the fold-out drawings can be avoided. Finally, the use of lines as picture primitives enables the partial occlusion of one object by another to be represented; in this drawing the radio is shown overlapping the box and the box is shown overlapping the far edge of the table. In this particular case, however, poor planning has resulted in a transparency and there is a false attachment between the line representing the right hand edge of the box and a line on the surface of the radio. As

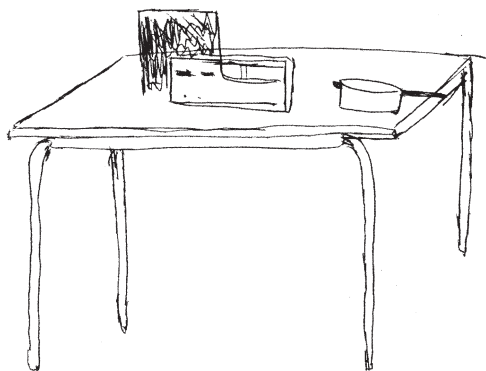


Figure 10. An example of a table with various objects on it drawn from a view of a real scene (Willats 1977a). This drawing is in oblique projection, with lines as picture primitives representing edges. The mean age of children producing drawings of this type was 13 years 6 months. Taken from Willats (1977b, Figure 6). Collection of the author.

a result the only true T-junctions representing points of occlusion in this drawing occur where the front edge of the tabletop overlaps one of the far legs. In developmentally later drawings of the same view of this scene more true T-junctions representing points of occlusion between the box, the radio and the far edge of the table were found to be present (Willats 1977a).

The drawing, denotation and mark systems children use in their drawings thus follow a definite developmental sequence. Using Arnheim's terminology, some of these changes can be thought of as changes in the graphic equivalents that children use at different stages. To take an example used by Arnheim one such graphic equivalent concerns what I have called the representation of extendedness. In the tadpole figures, for example, round areas on the picture surface are used to represent round objects such as the head or head/body, while long areas or lines are used to represent long objects such as the arms and legs, and this kind of graphic equivalent persists in children's drawings for a long time. However, the simple term 'graphic equivalent' is less helpful in describing the complex changes that take place in the structural systems used by older children. Nor can this approach explain the characteristic anomalies that appear at different developmental stages, such as the fold-out drawing of the box standing on the tabletop in Figure 8 or the transparency in Figure 10.

Using the terminology developed in later work in vision science and artificial intelligence thus provides a way of enhancing Arnheim's concept of graphic equivalents and the role they play in children's drawings. Another advantage of using this terminology is that a more explicit account can be given of the structural systems used by the artists of the avant-garde, and the similarities between these structural systems and those used by young children.

These similarities have often been remarked and are not coincidental. The beginning of the twentieth century saw an explosion of interest in children's drawings, not just on the part of psychologists and teachers but also among artists, and children's pictures were often displayed alongside the work of the avant-garde. Franz Cizek pioneered the belief in children as naturally creative artists, and in 1908 work by children attending his art classes was exhibited at the Kunstschau in Vienna. In Germany, children's pictures were reproduced alongside paintings by expressionist artists in *Der Blau Reiter*, edited by Kandinsky and Franz Marc, and in Italy the Futurist Exhibition in Milan in 1911 also included children's drawings. When Marion Richardson, a young art teacher at Dudley High School in England showed some of her children's work to Roger Fry in 1917, Fry added them to his current Post-Impressionist Exhibition, which was dominated by Picasso and Matisse (Malvern, unpublished 1988; 1995). In Russia children's drawings were exhibited together with the work of avant-garde artists in successive years from 1908 to 1911, and in 1913 (Molok 1998).

Thus by the early years of the twentieth century many European artists of the avant-garde had seen children's paintings and drawings exhibited together with their own paintings and drawings. Many of these artists incorporated motifs taken from children's drawings and paintings in their own work, and some of them, including Klee and Kandinsky, made substantial collections of pictures by children (Fineberg 1997). What was it about these pictures that attracted their interest?

There is no single, simple answer to this question, but one of the reasons why so many avant-garde artists became interested in children's drawings was that they saw in children's drawings many of the elements they needed in their search for a new pictorial language. Arnheim put it this way:

To break with a tradition that has run its course and to reinvent the world of imagery, artists tend to look around for models. The guidance and inspiration they derive from remote sources demonstrates that productive help can be obtained from communication that is at best partial. Just as European painters and sculptors received a needed impulse from African carvings, about whose meaning and function they knew next to nothing, so the strong influence of children's drawings relied on precepts, interpretations and connotations that had little to do with the states of mind producing these unassuming pictures. All that was needed were some of the formal properties for which artists were searching to resolve problems of their own. (Arnheim 1998: 15)

What were these 'problems'? And what were the 'formal properties' in children's drawings that the artists of the avant-garde recognised, often at an unconscious level, as a way of solving them?

In his account of Cubism the critic Jacques Rivière, writing in 1912, described the problems that were currently being encountered by artists such as Picasso, Braque, and Léger. Cubism, Rivière believed, was an idea that was both interest-

ing and important, but the failure of these artists to solve the problems they were encountering was making them gloomy and insecure:

They are pioneers – and like all pioneers *maladroit* – of a new art that is now already inevitable. If they are so insecure, it is because the need that drives them insists on remaining implicit. It exploits them as if they were soldiers on manoeuvres, offered no reasons for what they are doing, and left to fend for themselves because they do not grasp the underlying point. . . My intention is to offer the cubists a little more freedom and confidence, by furnishing them with the fundamental reasons behind what they are doing. (Rivière 1912/1999:2)

Rivière goes on to say that the fundamental difference between the Cubists and the painters who preceded them was that they were attempting to paint the essence of things, rather than as they appear from a particular point of view. Ever since the Renaissance painting had been based on optics and the laws of perspective. But, Rivière argued, lighting and perspective alter and distort the true shapes of objects, so that in order to achieve their ends the Cubists would have to renounce both perspective and *chiaroscuro* lighting.²

However, if perspective and *chiaroscuro* lighting were to be abandoned they had to be replaced by other ways of representing shape and space. Rivière offered a variety of solutions to this problem but I shall describe only one of them, chosen because of its obvious relevance to children's drawings. 'The suppression of perspective gives rise to this simple rule: the object must always be presented in its most revealing orientation, it should be presented in the attitude where the object is most itself, so that we are able to see it well' (*ibid.*:8). This is virtually identical to the feature of children's drawings that Luquet (1927/2001) called 'exemplarity', in which each object is drawn so that its shape and essential features are most characteristic. In the case of rectangular objects this amounts to saying that each face should be drawn as a true shape.

Rivière then goes on to say that following this rule could 'sometimes even involve several viewpoints; sometimes an object will reveal itself in a way that it would be impossible for us ever to see it, with one face more than would be revealed to us from a stationary position' (*ibid.*:9). This is what Luquet called 'rabattment' in children's drawings, but this feature almost inevitably leads to what Rivière called 'the first error of the Cubists'; that is, to anomalies similar to children's fold-out drawings.³

Thus the application of rules that involve the representation of the faces of objects as true shapes on the part of the Cubists not only led them to use drawing systems similar to those used by children, but to the appearance in their work of the same characteristic anomalies. Developmentally, children either overcome this problem by restricting the number of faces to two, or by changing to a different drawing system such as oblique projection or some form of perspective. Cubist

painting, however, continued to contain anomalies of this kind and in the event proved to be a relatively short-lived movement.

Pictorial anomalies are usually regarded simply as errors to be eliminated, but they can also be used deliberately as a way of investigating the nature of depiction. This technique was used in the 1970s by a number of workers in the field of artificial intelligence in order to make explicit the rules of line drawing. Clowes (1971) and Huffman (1971), for example, described the rules governing the denotation systems in line drawings of objects with plane faces, and showed that breaking these rules resulted in pictures of 'impossible objects'. The main problem in analysing pictures was seen to be the problem of finding ways of describing the elements – lines, blobs, dots and so on – of which pictures are composed and the relationships between them. In other words, finding ways of describing what I have called the drawing systems and the denotation systems.

These studies were greatly influenced by Noam Chomsky's account of generative grammar, and his use of nonsense sentences such as 'colorless green ideas sleep furiously' as a way of exposing our intuitions about the rules of language is well known (Chomsky 1965). Long before this, however, a number of avant-garde artists seem have used pictorial anomalies deliberately as a way of investigating the nature of depiction. Among them were Juan Gris and Paul Klee.

Gris's *Breakfast*, 1914 almost amounts to a handbook on pictorial anomalies (Willats 1997).⁴ Some of these anomalies – the use of axonometric projection, the transparencies, the false attachments and the inclusion of text in the pictures – resemble those found in children's drawings, but there is no evidence (so far as I know) to suggest that Gris was interested in children's drawings or was influenced by them in his work. Gris wrote little about his methods, but in his attempts to represent the three-dimensional reality of objects on a two-dimensional surface without recourse to illusion it seems likely that he was influenced by engineering and architectural drawings. The use of axonometric projection for architectural drawings had been popularised by Auguste Choisy at the end of the nineteenth century, and the inclusion in this painting of parts of objects that would not otherwise be seen is not unlike the use of sectional views in engineering drawings. Most importantly, perhaps, the basic denotation system in this painting, one in which lines in the picture are used to denote edges and contours in the depicted scene, is the one commonly employed by engineers, even though in this case the normal rules are reversed.

Paul Klee, in contrast, used denotation systems that are quite unlike those used either by adult artists working in the Renaissance tradition, or by architects or engineers, and much more like those used by young children. At the time he was working comparisons were very often made between his paintings and drawings and those produced by children, and Klee was well aware of this:

Child's play! Those gentlemen the critics often say that my pictures resemble the scribbles and messes of children. I hope they do! The pictures that my little boy Felix paints are often better than mine because mine have often been filtered through the brain.
(cited in Wilson 1992: 18)

As Fineberg (1997) has shown, Klee used motifs taken from drawings by his son Felix, from paintings and drawings he himself produced as a child, and from pictures by other children. For example, his drawings *Railroad Engine*, 1939 and *Mountain Railroad*, 1939 are clearly influenced by a drawing of an engine and carriages produced by Felix in 1913. Drawings and paintings inspired by his own childhood drawing of a boat produced in 1889 include *The Road from Unklaitch to China*, 1920, *Memory of Lugano*, 1921, *Yellow Harbor*, 1921 and *Steamship before Lugano*, 1922. Finally, Klee's *Pastoral (Rhythms)*, 1927 looks remarkably like a child's drawing taken from the collection of Kandinsky and Münter. Fineberg (1997) gives numerous other examples.⁵

If this were all, we might think that Klee's work consisted of little more than clever pastiches of children's drawings. But whereas the anomalies in children's drawings are the unwanted by-products of the representational systems that they use at different developmental stages, Klee used similar anomalies quite deliberately, both for expressive reasons and as a way of investigating the nature of depiction.

Figure 11 shows Klee's *Shipwrecked*, 1938. The absence of T-junctions and end-junctions in this drawing suggests that it is based on a denotation system in which regions denote volumes, with the regions subject to the shape modifiers 'having smooth outlines', 'being pointed' and 'being bent'. In this respect the denotation system Klee used is similar to that in the drawing of a man shown in Figure 7, but there are two elements in Klee's drawing which prevent it from being a mere pastiche of a child's drawing. The first is the absence of the eyes, nose and mouth or other defining features, which are nearly always present in children's drawings. The second is that in the figure on the left in *Shipwrecked* the regions representing the head, one arm and one leg are not joined to the body as they would be in a child's drawing.⁶ Klee is thus experimenting with some of the representational systems that he would have found in children's drawings rather than merely replicating their surface appearance.

In his *Gentle Drumroll*, produced in the same year (Figure 12) Klee took his experiments with representational systems a stage further. Although the shape of the region representing the drum resembles that of the boat in *Shipwrecked* the outlines have been continued beyond the corners, so that it is difficult to interpret them as either as lines standing for contours or as the outlines of regions. The main effect of these extensions beyond the outline is to emphasise the status of these lines as marks, thus flattening the picture surface. The two oblique lines at the bottom left are ambiguous and can be interpreted either as drumsticks or as



Figure 11. Paul Klee *Shipwrecked*, 1938. Crayon, 29.7×20.3 cm. Klee-Stiftung, Kunstmuseum, Berne. Copyright DACS.

arms, and the role of the single vertical line is also unclear, although the lower half looks like part of the body with the arms attached to it. In contrast, the curved line running from top to bottom in the drawing clearly represents the contours of the hair, the head and the shoulder, and the lines in the middle represent the contours of the mouth and lips. In this drawing Klee thus combined at least two quite different denotation systems, one in which the lines stand for long volumes as they do in tadpole figures, and the other in which the lines stand for contours as they would do in an adult drawings. In children's drawings these systems appear at the beginning and end of the developmental sequence, but no child would combine such different systems in one drawing.

In the next year, in his *Childlike Again* 1939, Klee again combined two very different denotation systems. The large, roughly square area in the middle of the drawing is framed by four lines. The line on the left stands for the animal's lead as a thin, wire-like form. This would be a perfectly legitimate denotation system in an adult drawing, except that at its lower end the line turns into the contour of the underside of the animal's body. The horizontal line at the top stands for the upper parts of the figure's arms as volumes, as it might do in a tadpole figure, and the vertical lines stand for the lower arms, ending in small round regions representing the hands. The head (except that it is turned on one side) could also have come straight from a child's drawing. The lines defining the remaining two sides of the square, in contrast, appear to stand for contours. At the bottom the shape of the

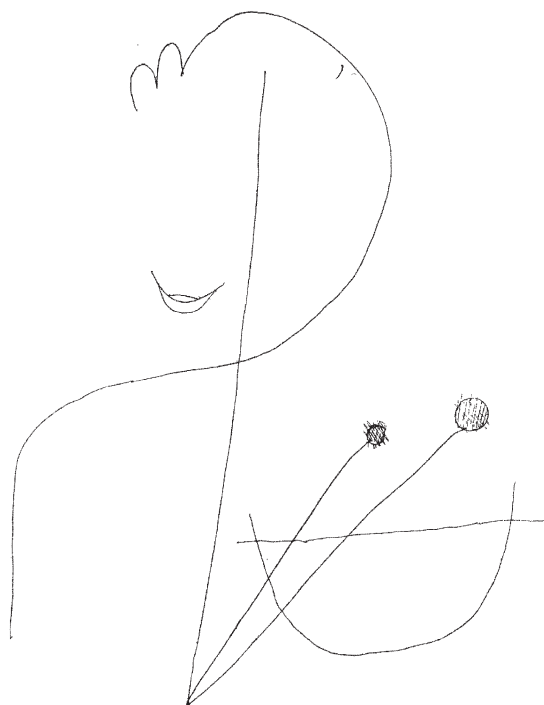


Figure 12. Paul Klee *Gentle Drum-roll*, 1938. Crayon, 27.1×21.3 cm. Klee-Stiftung, Kunstmuseum, Berne. Copyright DACS.

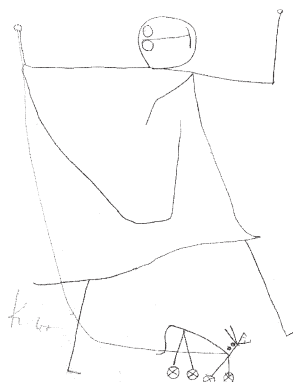


Figure 13. Paul Klee *Childlike Again*, 1939. Pencil on paper on linen backing, mounted on cardboard. Klee-Stiftung, Kunstmuseum, Berne. Copyright DACS.

line suggests the edge at the bottom of a dress, and the legs, which again could have come from a tadpole figure, nevertheless intersect this edge in a way that suggests that the T-junctions stand for points of occlusion. The shape of the line on the

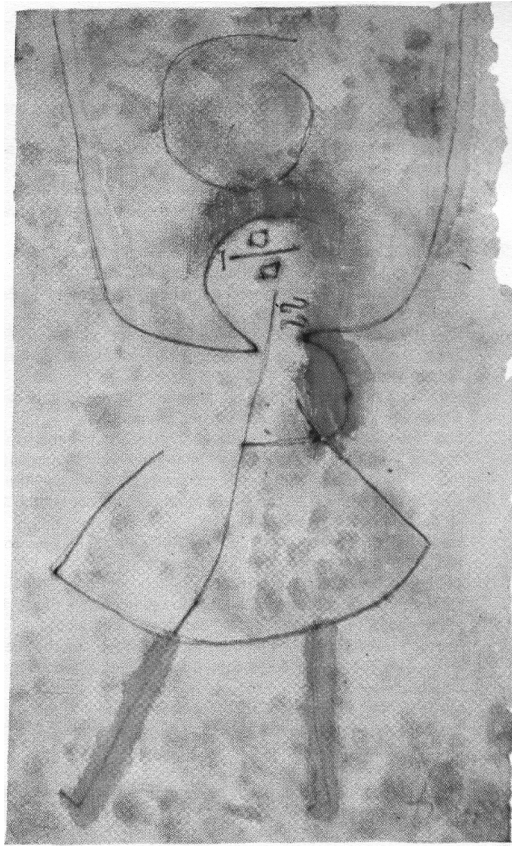


Figure 14. Paul Klee *With Green Stockings*, 1939. Watercolour on blottingpaper, 35.0×21.0 cm. Felix Klee Collection, Berne. Copyright DACS.

right also suggests that it represents a contour, and this reading is supported by the way in which the lines within the square end, suggesting end-junctions standing for points of occlusion within the drapery. Again, this anomalous mixing of denotation systems from opposite ends of the developmental sequence is something that would never occur in a child's drawing.

In these three drawings Klee seems to be finding his way, but in *With Green Stockings*, 1939 (Figure 14) he brought his experiments with the denotation systems to a triumphant conclusion. As in *Gentle Drumroll*, 1938 and *Childlike Again*, 1939 this painting is based on two different denotation systems: one in which the lines as picture primitives stand for volumes and the other in which they stand for edges and contours. In *With Green Stockings*, however, these two systems are not used independently but are combined in such a way that the lines change their meaning along their length.⁷

To begin at the top of the painting, the line on the left starts life by standing for the girl's arm as a volume, as it would do in a tadpole figure. As it turns, however, it begins to stand for the contour of the girl's shoulder and then, rising again, for the contours of the neck, chin and cheek. The corresponding line on the right also begins life as a volume, but then turns into the contour of the back of her bodice. Once past the contour it then changes again in order to represent the long flat area of a belt. The changes of meaning in the more or less vertical line in the centre of the painting are even more complex. At the bottom, single lines are used to stand for the foot and leg. Then as it passes the hem of the skirt at a false T-junction the line changes its meaning to represent the contour of a fold in the drapery of the skirt. As it passes the end of her belt the meaning becomes ambiguous; is this line still the contour of a fold in the drapery, or the bounding contour of her bodice? Finally, the end of this line stands for the ridge in the girl's neck that appears as she turns her head.

Two more major lines remain. The first stands, conventionally enough, for the contours of the skirt, except that Klee contrasts the true end-junction standing for a point of occlusion on the right with a false end-junction where the contour of the skirt ends in mid air on the left. The remaining line stands for the contours of the girl's ball, but even this is anomalous because the contours fail to join. Even very young children usually take great care to join the ends of their lines where they represent the outlines of regions denoting volumes, as in the round region representing the head/body in Figure 3.

'Art does not reproduce the visible,' Klee noted in 1918, 'rather, it makes visible' (cited in Fineberg 1997:90). Certainly, none of the lines in Klee's drawings I have discussed above reproduce the visible in any simple way, and the same is true of most children's drawings. Klee's remark could easily have been made by Arnheim, and for Arnheim these lines are not copies of features of the visible world, but instead provide graphic equivalents for these features in a given medium. However, we need to go further than this if we are to give a full account of children's drawings and Klee's drawings and paintings.

Drawings by young children are based on representational systems that are fundamentally different from those used by most adults, and look anomalous judged by the standards of academic realism. Anomalies of this kind include single lines standing for whole volumes, fold-out drawings of rectangular objects in which the edges and corners cannot be made to join up properly, and the transparencies arising from children's reluctance to draw only the parts of faces or regions that are in view rather than their true shapes. Once children become aware of these anomalies they try to eliminate them from their drawings, and their attempt to do this provides the mechanism of development (Willats 1984).

Many of Klee's drawings and paintings are, like children's early drawings, based on unfamiliar representational systems and, like children's drawings, look anoma-

lous to adult eyes. But as I have shown, many of his later works contain anomalies that would never appear in children's drawings. *With Green Stockings* provides a good example. The single lines representing the arms and legs could come straight from tadpole figures, but children would never change the meanings of these lines into edges and contours.

The anomalies in children's drawings occur accidentally, but Klee seems to have introduced them quite deliberately. As he said in his diaries: 'I have carried out many experiments with laws and taken them as a foundation. But an artistic step is taken only when a complication arises' (Klee 1961:454). Klee wrote a great deal about his work, but perhaps the strongest evidence that Klee was using the anomalies in *With Green Stockings* deliberately lies in the painting itself. The use of colour in this painting, apparently only decorative, is in fact crucial to its meaning. At precisely the points where the lines representing the arms and legs change their meaning these changes are marked by the edges of patches of colour: yellow for the arms and green for the legs. Hence the title of the painting: *With Green Stockings*. The media Klee used for this painting, watercolour on blotting paper, would have been thoroughly despised by the academic painters who preceded him, the handling is not particularly skilful, and as I have shown this painting shares some of the formal characteristics of drawings by very young children. What makes this a major work of art, however, is the clarity and precision of Klee's visual thinking.

Notes

1. Gibson (1978) also rejected the copying theory of depiction. He suggested instead that pictures provide an array of graphic invariants that correspond to the invariants in the array of light coming from the scene. As examples of invariants Gibson gave a list of what are now usually referred to as nonaccidental properties such as straightness, curvature, parallelism and closure (Biederman 1987). Gibson's graphic invariants are not unlike Arnheim's graphic equivalents, but unlike Arnheim Gibson never attempted to test his theory by analysing either children's drawings or artist's pictures in any detail.
2. 'A house which has trees put in front of it by perspective can end up as two, disconnected, white triangles. A river, seen through a row of poplars, can appear as a string of small brilliant diamond shapes. Yet is not the essence of a house that it is a cube, and of a river that it flows in a single, deep and continuous mass? It is true that reality shows us objects mutilated in this way. But in reality we can move around: a step to the right and a step to the left completes our vision. Our knowledge of an object is, as we have said, a complex sum of perceptions. Since the plastic image does not move, the plastic image must be complete within a single glance and so has to reject perspective' (Rivière 1912/1999:8).
3. 'From the fact that the painter should show enough faces of the object to suggest its volume, they conclude that they ought to show all of the faces. From the fact that it is sometimes necessary to add to the visible faces one which could not be seen without displacement, they conclude that they have to add all of those which could be seen by moving around the object and inspect-

ing it above and underneath. There is no need to dwell on the absurdity of such inferences. We will simply remark that the process, as it is understood by the cubists, leads to a result quite counter to that which they try to obtain. If a painter sometimes shows more faces of the object that one can see at once in reality, it is in order to give it volume. But the whole volume is closed and implies the return of those surfaces upon themselves; the volume consists of a certain relation of all the faces to a centre. In putting all the faces beside one another, the cubists make the object seem like a folded-out card and they destroy its volume' (Rivière 1912/1999: 14, 15).

4. Gris's *Breakfast*, 1914 is reproduced in colour and analysed in detail in Willats (1997: 275–279, colour plate 6).

5. Klee's *Railroad Engine*, 1939, *Mountain Railroad*, 1939, *The Road from Unklaitch to China*, 1920, *Memory of Lugano*, 1921, *Steamship before Lugano*, 1922, *Yellow Harbor*, 1921 and *Pastorale (Rhythms)*, 1927 are reproduced in Fineberg (1997) as Figures 4.47, 4.49, 4.35, 4.36, 4.37, 4.40 and 4.18.

6. In this respect, the shapes of the regions and the relations between them in Klee's drawing resemble those in Matisse's late *gouaches découpées*, although it precedes them by at least ten years.

7. Klee's *With Green Stockings*, 1939 is analysed in greater detail in Willats (2003).

References

- Arnheim, R. (1956). *Art and Visual Perception: A Psychology of the Creative Eye*. London: Faber and Faber.
- Arnheim, R. (1974). *Art and Visual Perception: A Psychology of the Creative Eye. The New Version*. Berkely and Los Angeles: University of California Press.
- Arnheim, R. (1998). Beginning with the child. In J. Fineberg (Ed.), *Discovering Child Art: Essays on Childhood, Primitivism and Modernism* (pp. 15–26). Princeton: Princeton University Press.
- Biederman, I. (1987). Recognition by components: A theory of human image understanding. *Psychological Review*, 94, 115–147.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Clark, A. B. (1897). The child's attitude towards perspective problems. *Studies in Education*, July, 283–294.
- Clowes, M. B. (1971). On seeing things. *Artificial Intelligence*, 2 (1), 79–116.
- Fineberg, J. (1997). *The Innocent Eye: Children's Art and the Modern Artist*. Princeton, NJ: Princeton University Press.
- Gibson, J. J. (1978). The ecological approach to the visual perception of pictures. *Leonardo*, 11, 227–235.
- Huffman, D. A. (1971). Impossible objects as nonsense sentences. In B. Meltzer & D. Mitchie (Eds.), *Machine Intelligence*, Vol. 6 (pp. 295–323). Edinburgh: Edinburgh University Press.
- Hollerbach, J. M. (1975). Hierarchical shape description of objects by selection and modification of prototypes. MIT Artificial Intelligence Laboratory Technical Report No. 346 (pp1–237). Cambridge, MA: MIT Press.
- Kemp, M. (1992). *The Science of Art: Optical Themes in Western Art from Brunelleschi to Seurat*. New Haven and London: Yale University Press.
- Klee, P. (1961). *Notebooks (The Thinking Eye)*. Vol. 1. (J. Spiller, ed.). London: Lund Humphries.

- Luquet, G.-H. (1927/2001). *Le Dessin Enfantin*. Paris: Alcan. Trans. Costall (2001). London: Free Association Books.
- Malvern, S. (1988, unpublished). Modernism and the child. Innovation in art and art education and the teaching of Marion Richardson and Franz Cizek.
- Malvern, S. (1995). Inventing 'child art': Franz Cizek and modernism. *British Journal of Aesthetics*, 35 (3), 262–272.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W. H. Freeman.
- Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London, Series B*, 200, 269–294.
- Matthews, J. (1999). *The Art of Childhood and Adolescence*. London: Falmer Press.
- Molok, Y. (1998). Children's drawings in Russian Futurism. In J. Fineberg (Ed.), *Discovering Child Art: Essays on Childhood, Primitivism and Modernism* (pp. 55–67). Princeton: Princeton University Press.
- Nicholls, A. L., & Kennedy, J. (1992). Drawing development: From similarity of features to direction. *Child Development*, 63, 227–241.
- Rivière, J. (1912/1999). Sur les tendances actuelles de la peinture. *Revue d'Europe et d'Amérique*, 1, 383–406. Trans. Costall, A. (1999) unpublished.
- Sörög, G. (2002). The classical concept of mimesis. In P. Smith & C. Wilde (Eds.), *A Companion to Art Theory* (pp. 19–28). Oxford: Blackwell.
- Willats, J. (1977a). How children learn to draw realistic pictures. *Quarterly Journal of Experimental Psychology*, 29, 367–382.
- Willats, J. (1977b). How children learn to represent three-dimensional space in drawings. In G. Butterworth (Ed.), *The Child's Representation of the World* (pp. 189–202). London: Plenum Press.
- Willats, J. (1984). Getting the drawing to look right as well as to be right: The interaction between production and perception as a mechanism of development. In W. R. Crozier & A. J. Chapman (Eds.), *Cognitive Processes in the Perception of Art* (pp. 111–125). Amsterdam: North-Holland.
- Willats, J. (1997). *Art and Representation: New Principles in the Analysis of Pictures*. Princeton: Princeton University Press.
- Willats, J. (2003). Optical laws or symbolic rules? The dual nature of pictorial systems. In H. Hecht, R. Schwartz, & M. Atherton (Eds.), *Looking into Pictures: an Interdisciplinary Approach to Pictorial Space*. Cambridge, MA: MIT Press.
- Willats, J. (in press). *Making Sense of Children's Drawings*. London: Lawrence Erlbaum Associates.
- Wilson, B. (1992). Primitivism, the avant-garde and the art of little children. In D. Thistlewood (Ed.), *Drawing Research and Development* (pp. 14–25). Harlow, Essex: Longmans.

Visual perception and theories of painting

An uneasy complementarity

Alf C. Zimmer

Introduction: What can psychological theories of perception contribute to theories of painting?

In the world of art the pervasive opinion concerning this question is “few if anything at all” and the few which can be learned is confined to the aesthetics of reception. There are rare exceptions to be found; e.g. Gibson’s and Gombrich’s discourse in “Leonardo” 1978/9 might be regarded as the most prominent example of a fruitful exchange between perceptual science and theory of art. In this discourse Gibson shows that under natural conditions the traditional distinction between appearance and reality does not hold. Under this condition of – what he calls – “direct perception”, the perceiver is not confined in his act of perceiving to a multitude of snapshot views (appearances) from which one has to infer the “reality” of the scene, but directly picks up the invariances of the scene and thereby has direct access to – what Gibson called – the affordances of the scene. The affordances (all the possibilities a scene *affords* the perceiver to act upon) makes up the “reality of the scene” for the perceiver. These affordances are not inferred (consciously or unconsciously) but are immediately given in the perceptual system.

If – as Friedrich Schiller analyses in his “Briefe über die ästhetische Erziehung des Menschen” (Letters on the aesthetic education of man, 1794/95) – the traditional task of painting in our culture is the imitation of the appearances (“Schein” or mimesis) then pictures can not convey reality in the way direct perception does under natural conditions. Therefore – according to Gibson – the perception of pictures is indirect perception, necessitating inferences (as postulated Helmholtz for any kind of higher perception) in order to deduce the reality *represented* in the picture. For Schiller, however, the understanding of pictures is not primarily a cognitive act but results from the human faculty to play (letter 15: “Der Mensch. . . ist nur da ganz Mensch, wo er spielt”). If painting is mimesis, then the understanding of painting is play.

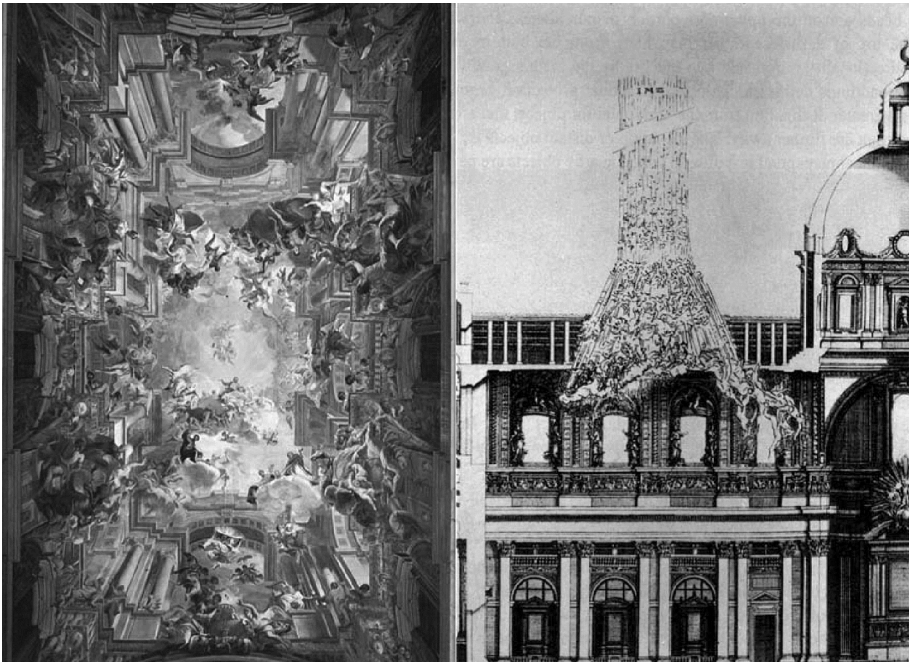


Figure 1. The ceiling of S. Ignacio as seen and the intended impression.

A striking example for this is the painting on the ceiling of S. Ignacio in Rom where the depicted space (see Fig. 1) can only be “seen” as the intended mimesis, if the viewer stands on a specific point in the church.

In the following I want to show that not only the psychology of perception can contribute something to a theory of art (Gombrich has done this already, see especially 1972) but that Gibson’s skepticism about the information contained in pictures is only partially valid. Furthermore I want to claim that closer analyses of painters’ implicit theories of perception can contribute something about the topical discussion in the 15th d 16th century on the theory of painting, namely, “Disegno” vs. “Colore” or the “Southern” and the “Northern mode of seing”.

Starting example: Reality in pictures

We all know that contrary to the legend told about Zeuxis birds do not get fooled by painted fruit because – as Gibson would point out – painted fruit does not exhibit the invariances of real fruit and therefore does not provide the affordances for the birds. However, more detailed analyses of phenomena show that this answer

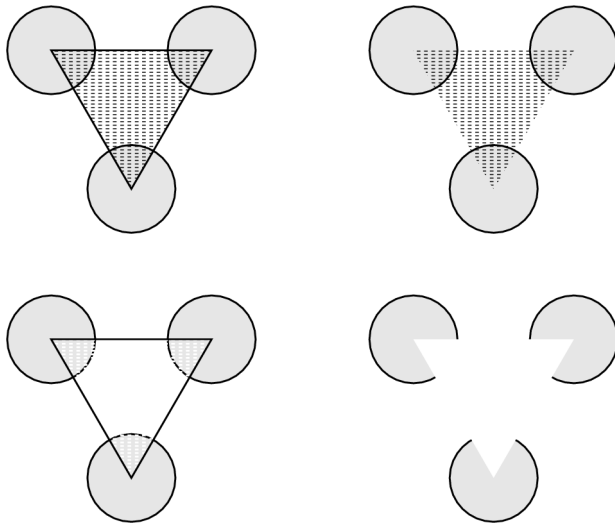


Figure 2. The relation between apparent transparency and illusory contours.

underestimates the information provided by pictures as a simple demonstration can show (see Fig. 2).

Simply by changing the greyness values in the different segments systematically, at first phenomenal transparency appears and finally even illusory contours (see Kanizsa 1979). For painters the techniques for producing these effects are part of the craft since antiquity (see the still life of a glass bowl with fruit in the villa of Julia Felix in Pompeii). However, experimental psychology can show that for the perceiver there is more in such pictures than an imitation of appearances: The perceived distance to the transparent layer is less than to the background and the figure constituted by the illusory contours is hovering above the plane of the circles and their surrounding. That is, the picture conveys – at least partially – affordances, albeit illusory ones: they not only fool the eye but also the pointing finger.

Why theories of perception tend to be neglected in the teaching of painting

The inherent tension between the art of painting and theories of perception becomes pointedly parent in the Bauhaus curriculum which was meant to comprise all preconditions for the craft which enables to create art that in itself cannot be taught. For didactic and theoretical purposes the authors of the curriculum distinguish between the faculty of creating art and the skills (craft) as its necessary but not sufficient prerequisite, that is, between art and craft. This curriculum –

insofar as it is concerned with the *craft* of painting – reduces perception, especially color perception, to optics; therefore physics is part of the curriculum and chemistry as far as necessary for dyeing and color production. Due to the fact that the creative act is regarded as beyond teaching, scientific approaches to art perception are not included in the curriculum: neither Helmholtz' physiological nor Hering's or Mach's phenomenological approaches. Despite of the fact that in the 1920s the concept of Gestalt dominated the discussion between psychologists, epistemologists and physicists in Central Europe and that the leading figures in Dessau or Weimar were well aware of the concept, they did not regard it as contributing to the craft and therefore as a topic for the curriculum.

Actually, the Bauhaus position can be seen as an attempt to circumvent the controversial discussion in the Renaissance and Baroque time about the proper theory of painting, namely, what Nicolas Poussin in his letter to De Noyers called "deux manières de voir les objets": "Two ways of viewing objects: Simply seeing them, and looking them attentively. Simply seeing is merely to let the eye take in naturally the form and likeness of the thing seen. But to contemplate an object signifies that one seeks diligently the means by which to know the object well, beyond the simple and natural reception of its form in the eye. Thus it can be said that mere *Aspect* is a natural operation, and that what I call *Prospect* is an office of reason which depends on three things: *The discriminating eye, the visual array and the distance from the eye to the object*" (Jouanny 1911: 143, my italics). These different ways of seeing reflect the conflict between "painting what is seen" vs. "painting what is known"; or more technically: Seeing an object from a fixed viewing point in contrast to perceiving an object as it is – namely including all possible transformations, that is its invariances (Gibson).

Actually, Michelangelo (as quoted by Francesco de Hollanda) contrasts Italian vs. Flemish painting in a similar way: "In Flanders they paint with a view to external exactness or such things as may cheer you and of which you cannot speak ill, as for example, saints and prophets. They paint stuffs and masonry, the green grass of the fields, the shadow of trees, and rivers and bridges, which they call landscapes, with many figures on this side and many figures on that. And all this, so it pleases some persons, is done *without reason or art, without symmetry or proportion, without skill for choice or boldness and, finally without substance or vigour* (de Hollanda 1928: 16).

In a certain way, the differences between the Southern and the Northern view of painting can be traced back to different interpretations of Alberti's treatise on painting (1972: 53) "Large, small, short, long, high, low, wide, light, dark, bright, gloomy and everything of the kind, which philosophers termed accidents, because they may or may not present in things, – all these are such as to be known only by comparison – . . . as man is best known of all things to man, perhaps Protagoras, in saying that man is the scale and the measure of all things, meant that accidents in

all things are duly compared to and known by the accidents in man.” What Alberti describes here is the concept of frame of reference in a general way. The different interpretations of “seeing” start with the meaning of “man is the scale and the measure of all things”, namely, if scale and measure are to be interpreted literally as in Poussin’s *Aspect*, that is, as in geometry of physics, or in a more general way, taking into account world knowledge and implicit cultural background as in Poussin’s *prospect*.

The Northern mode (aspect) is best represented by Johann Keplers theory of perception (Kepler, J. “Ad Vitellionem”, 143):

Thus vision is brought about by a picture (*pictura*) of the thing seen being formed on the concave surface of the retina. . . and thus the origin of errors in vision must be sought in the conformation and the functions of the eye itself.

From this follows the principle of the Northern mode “*ut pictura, ita visio*” and for this reason it is not astonishing that historians of art (e.g. Panofsky or Lord Clark) have interpreted the art of Jan van Eyck or Vermeer by using terms from photography: “Jan van Eyck’s eye operates as a microscope and a telescope at the same time” (Panofsky 1953, 1: 182) or about the View of Delft “this unique work of art is certainly the nearest which painting has ever come to a colored photograph” (Clark 1976: 263).

The camera metaphor for visual perception

The theory of vision as proposed by Kepler regards seeing as a passive process and Panofsky and Clark in using the metaphor of a photographic camera concur with Kepler “. . . the retina is painted with the colored rays of visible things.” According to this theory, the starting point for any perception is the projection of the external world upon the concave retina, actually Jan van Eyck alludes to this theory in “The Arnolfini Wedding” by showing the backs of the bride and the groom plus himself in a blue gown reflected on a convex mirror, thus representing the projection on the concave retina (Figure 3).

In this modeling of the perceptual process neither the movements of the perceiver’s eyes nor the observer’s movements relative to the perceived object are taken into account. Actually instruments like the Dürer window (see Figure 4) enforce the rigid constellation of percept and perceiver – and thereby making perspective painting easy.

Another mechanistic approach to perspective is the Camera Obscura which in the version of Drebbel has had a strong influence on the practice of painting in the Netherlands during the 16th and 17th century (see Figure 5). C. Huygens (1596–1687) who described in his autobiography of 1629 many details concerning



Figure 3. Detail of the Arnolfini Wedding.



Figure 4. The Dürer Window from 'Unterweysung der Messung', reflected for better comparability with Figure 9.

the cultural background of pictorial art in the Netherlands writes about Drebbel's instrument in a letter to his parents: "I have in my home Drebbel's other instrument, which certainly produces admirable effects in reflection painting in a dark room. It is not possible to describe for you the beauty of it in words: all painting is dead in comparison, for here is life itself, or something more noble, if only it did not lack words. Figure, contour, and movement come together naturally therein, in a way that is altogether pleasing" (quoted by Alpers 1983: 12).

There is strong evidence that the camera obscura was a common instrument for Flemish painters; Steadman (2001) even suggests that in Vermeer's studio was installed a huge camera obscura (see Figure 6).

In the 19th century Helmholtz proposed a technique for the demonstration of his theory regarding 'unconscious inferences' in perception by using a peep hole box as developed by van Hoogstraten (Perspectifkas, National Gallery London)

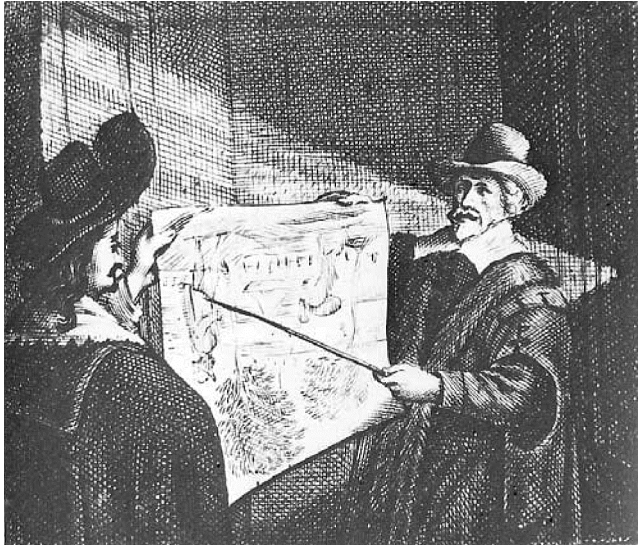


Figure 5. Johan van Beverwyck's (1667) theory of perception for artists in *Wercken der Genees Konste*.

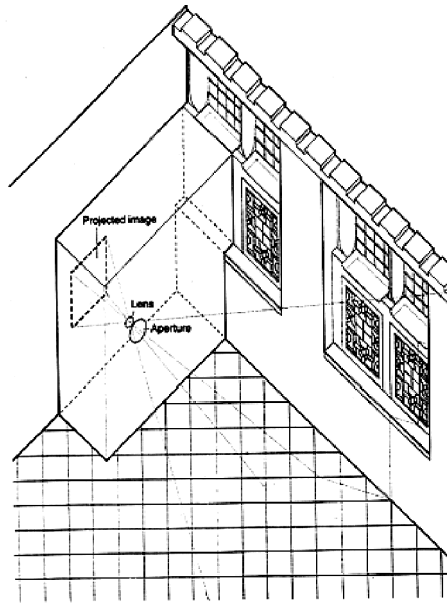


Figure 6. The lay-out of Vermeer's studio according to Steadman (2001). The windows and the tiling of the floor can be used to identify the position of projected images for the different interiors as painted by Vermeer.

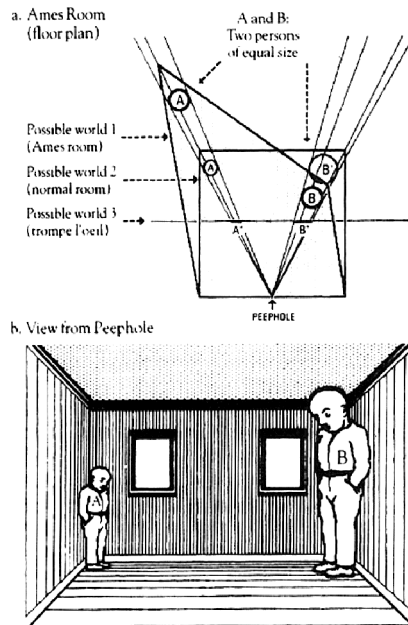


Figure 7. (a) the ground plan of the Ames room (b) the resulting appearance if viewed through a peep hole (after Shepard 1990).

but distorted in such a way that size illusion even with common objects result. This technique has resulted in the so-called Ames room (see Figure 7).

Helmholtz too regarded the process of *seeing* as passive as, that is, governed by the sensory input (Kepler's projections on the retina) but he proposed that higher cognitive processes lead to the conscious *perception*. In the case of the Ames room, the higher cognitive processes *interpret* the angles as perpendicular and therefore the observer *rescales* the size of the persons watched through the peep hole.

Apparently, Gibson had this in mind when he classified picture perception as well as Ames room perception as *indirect*, namely, either not affording the observer with invariant information or even inducing illusions. However, a closer look at the artists' work even if supported by the Dürer Window of the camera obscura shows that the artists intended to provide invariant information albeit not by means of perspective geometry.

How to depict the "what" and the "how"

If one reconstructs what the artist in Dürer's illustration really could have seen, one arrived at a grossly distorted reclining woman (see Figure 8).

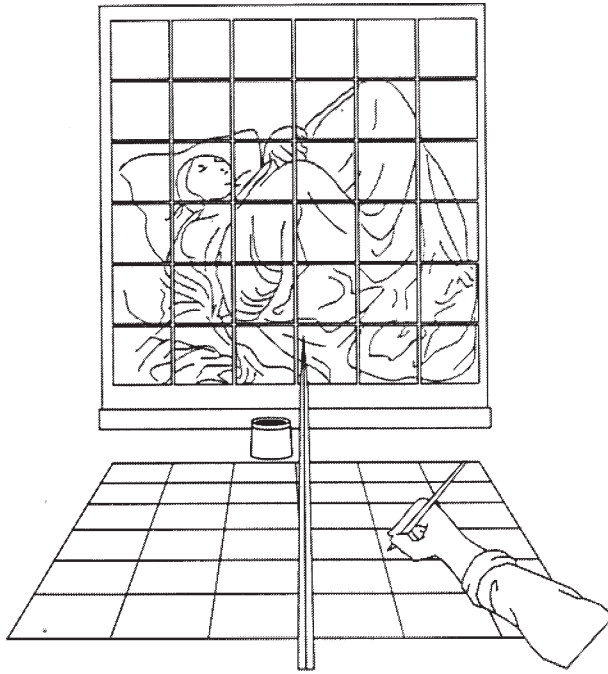


Figure 8. What the artist in Dürer's 'Unterweysung' saw (after Kubovy 1986).

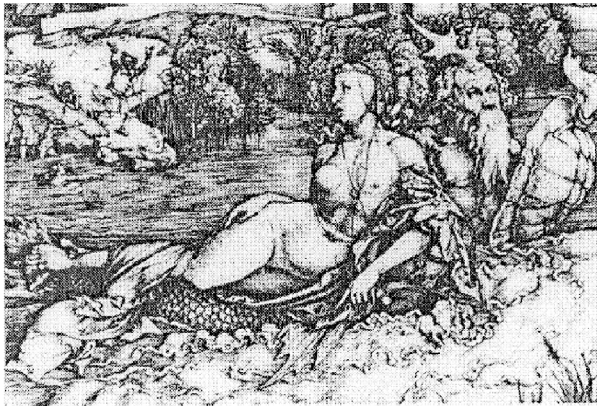


Figure 9. A reclining nude (Das Meerwunder 1498) by Dürer.

If one uses Dürer's graphic work for comparison, one can immediately see what for Dürer the artist has to do *beyond* following the rules in 'Unterweysung der Messung', namely, to stage the to be depicted objects in such a way that their major axes of orientation are parallel to the fronto-parallel plane of the observer (the grid in the Figure 8 above).

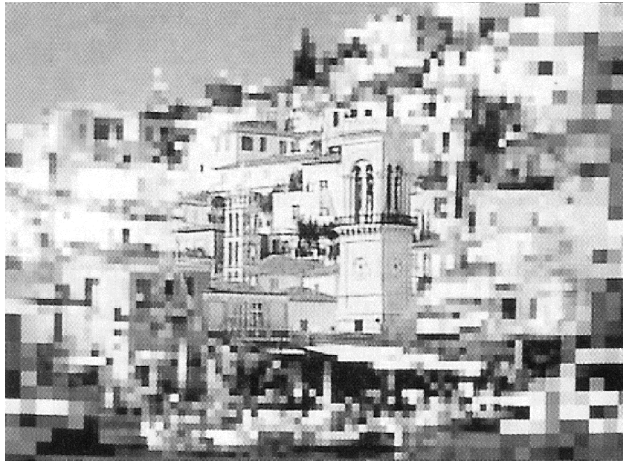


Figure 10. A demonstration of the degradation of acuity in the periphery of a glance. Actually, the saturation of the colors decreases, too.

The *staging* of the to be depicted objects combined with a *scaffolding* of the space into which they are embedded has been identified by Alpers (1983) as one of the core principles in Flemish painting.

However, his staging and scaffolding techniques can not explain the difference between the Northern and the Southern View in art because it is ubiquitous in paintings of this time. What is discriminating between these approaches can be seen in Dürer's *Unterweysung* where he uses a combination of two different perspectives: one for the artist and another for the reclining woman. And, actually, this combination of multiple vantage points is closer to the modern psychological view on the perceptual processes than Kepler's theory: The percept is the result of an integration of many glances at a scene, these different glances focus on different parts of the scene, directed by attentional processes, and their acuity is only high in the focus and then degenerates rapidly (see Figure 10).

Due to the blurriness in the periphery of the glances, it is possible that local deviations from the global orientation are not detected if a consistent global percept is possible, most of the so-called Gestalt phenomena work on the local as well as on the global level.

For the artist this implies that by mixing perspectives it is possible to depict the isolated object in its canonical form and at the same time to construct a global scene which has not to be consistent with the isolated views. For example, in Raphael's *School of Athens* Euclid and Ptolemy are depicted presenting perfect spheres, despite the fact that in a peep hole view of the general perspective these spheres had to be transformed into ellipsoids (see Figure 11).



Figure 11. Euclid and Ptolemy presenting spheres which are distorted according to the global perspective.

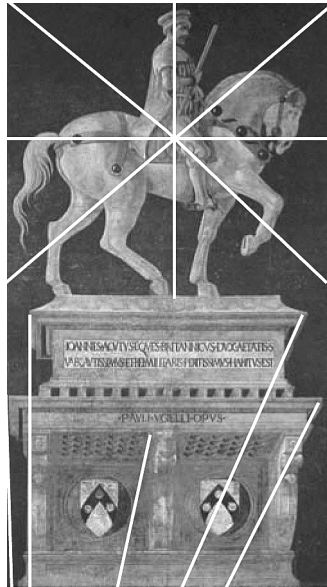


Figure 12. The epitaph of John Hawkwood with the respective lines converging to the different vanishing points.

By using this mix of perspectives it is possible to depict at the same time ‘what is known’, namely the sphericity, and ‘what is seen’, namely its position at unique point in the spatial scene. That painters like Raphael were well aware of this Platonic influence on their art becomes apparent in his letter to Baldassare Castiglione “... in order to paint one beautiful woman I’d have to see several beautiful women. ... I make use of a certain idea which comes to my mind. Whether it carries any excellence of art I do not know but I work hard to achieve it.”

Perhaps, this is the essence of the Southern view as implied in Michelangelo's critique of Flemish painting: "skilful choice or boldness .. substance or vigour." Uccello's painted epitaph of John Hawkwood exemplifies the application of these principles (see Figure 12).

At least partially, these examples show that in a kind of cooperation between the artist who depicts multiple views in order to show 'what we know' and the perceptual processes of the observer there emerge affordances for the perceiver about the world represented in the pictures.

Recent results from cognitive neurology (Zeki 2000) indicate that there is a neurological basis for abstraction which resolves the clashes between the sensory input and 'ideals'. Insofar the Southern Mode of depicting is in accordance with psychological approaches to picture perception. The interaction between perception science and pictorial art becomes obvious in Hockney's technique of assembling Polaroid shots made from different viewing points in such a way that the viewer is afforded a visual walkabout through the scene (e.g. *The Zen Garden*). This resembles Hochberg's (1962) theory of glance integration, insofar not only on a theoretical level as in Gombrich's and Gibson's discourse but also in the practice of art a level of complementarity between art and science can be achieved.

References

- Alberti, L. B. (1972). *On Painting and Sculpture* (ed. and transl. by C. Grayson). London: Phaidon Press.
- Alpers, S. (1983). *The Art of Describing*. Chicago: The University of Chicago Press.
- Clark, *Landscape into Art* (1976 new ed.). New York: Harper and Row.
- Gibson, J. J. (1978). The ecological approach to the visual perception of pictures. *Leonardo*, XI, 227.
- Gombrich, E. H. (1972). The "what" and the "how": Perspective representation and the phenomenal world. In R. S. Rudner & I. Scheffler (Eds.), *Logic and Art*. New York.
- Gombrich, E. H. (1979). Comment on Gibson (1978). *Leonardo*, XII, 174–175.
- Hochberg, J. (1962). The psycho-physics of pictorial perception. *Audio-visual Communication Review*, 10–50.
- Hollanda, F. de (1928). *Four Dialogues on Painting* (transl. by A. F. G. Bell). London: Oxford University Press.
- Jouanny, C. (Ed.). (1911). *Correspondance de Nicolas Poussin*. Paris.
- Kanisza, G. (1979). *Organization in Vision*. New York: Praeger.
- Kubovy, M. (1986). *The Psychology of Perspective and Renaissance Art*. Cambridge: Cambridge University Press.
- Panofsky, E. (1953). *Early Netherlandish Paintings*. Cambridge: Harvard University Press.
- Shepard, R. (1990). *Mindsights*. San Francisco: Freeman.
- Steadman, P. (2001). *Vermeer's Camera*. Oxford, New York: Oxford University Press.
- Zeki, S. (2000). Abstraction and idealism. *Nature*, 404, 547.

PART III

Bridging perception and depiction of visual spaces

Dynamics of picture viewing and picture description

Jana Holsanova

Introduction

In his book *Visual thinking*, Arnheim (1969) writes: "... cognitive operations called thinking are not the privilege of mental processes above and beyond perception but the essential ingredients of perception itself. I am referring to such operations as active exploration, selection, grasping of essentials, simplification, abstraction, analysis and synthesis, completion, correction, comparison, problem solving, as well as combining, separating, putting in context" (Arnheim 1969:13). He draws the conclusion that "visual perception (...) is not a passive recording of stimulus material but an active concern of the mind" (Arnheim 1969:37). In order to understand the dynamics of the ongoing perception and the cognitive processes involved, I investigate the dynamics of picture viewing and picture description (Holsanova 2001). I connect two disciplines, linguistics and vision research, and combine two methods, spoken language descriptions and eye movement protocols, to cast light on the underlying mental processes.

In the first section of this chapter, I present the data, method and analytical tools of the experimental studies. In the second section, I discuss the results of the empirical studies within the functionalist and information–processing framework. The third section addresses how the observers verbalize their impressions from the picture and how their picture descriptions differ in style. In the fourth section, I consider how the results from the studies contribute to an understanding of the dynamics of the ongoing perception processes – an area on the intersection of artistic and cognitive theories. Finally, in Section 5, I present an eye tracking study on picture viewing and mental imagery. The results are summarized and discussed in Section 6.

Picture viewing and picture description: Data, method and analytic tools in the experimental studies

I am going to present results of empirical studies where informants under different conditions (on-line or from memory) and in different settings (monological or interactive) have described a complex picture. In these studies, cognitively oriented research is used with the expectation that the analysis of eye movement data and spoken language description data can give us hints about the dynamics of the underlying cognitive processes. The trouble is, of course, that we cannot directly uncover the content of our mind. If we want to learn about it, we have to do that indirectly, via spoken language in action (via language's overt manifestations) and via the process of visual focusing (via fixations as overt visual manifestations). The closest we can come to throwing light on the mind is via the verbal and visual foci (see Figure 1).

On the one hand, eye movements reflect human thought processes. It is easy to determine which elements attract the observer's eye, in what order and how often. In short, eye movements offer us a window on the mind. On the other hand, verbal foci formulated during picture description are the linguistic expressions of a conscious focus of attention. With the help of a segmented transcript, we can learn what is in the verbal focus at a given time. In short, spoken language description offers us another window on the mind. Both kinds of data will be used as an indirect source to gain insights about the underlying cognitive processes and about human information processing.

In the first eye tracking study, informants were asked to verbally describe a complex picture which they were simultaneously watching (see Figure 2). This picture contains both states and events, allows different ways and scan paths and thereby also different 'readings'. The process of the visual discovery of the picture was registered using an eye tracker. The spoken language description was

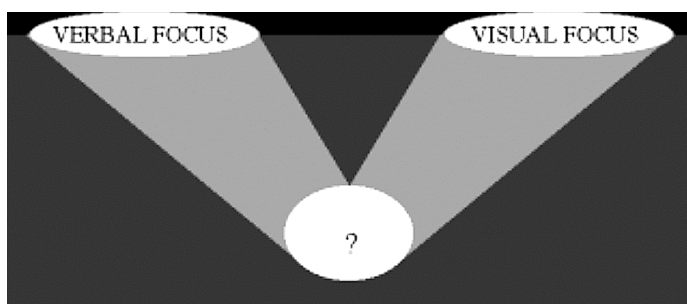


Figure 1. Verbal and visual focus.



Figure 2. The complex picture. The motif is from Nordqvist (1990).

recorded, transcribed and segmented into verbal foci and superfoci (Chafe 1980; Holsanova 2001).

What we attend to during the visual perception and the spoken language description can be conceived of with the help of a spotlight metaphor which provides intuitively a notion of limitation and focus: The picture elements fixated visually lie in the focus of a spotlight and are embedded in a context. The spotlight moves to the next area that pops up from the periphery and will lie in the focus for a while. If we focus our concentration and eye movements on a point, we mostly also divert our attention to that point. By using a visual fixation analysis, we can thus follow along the path of attention deployed by the observer. Concerning spoken language description, it has been shown that we focus on one idea a time (Chafe 1980, 1994). The picture elements described lie in the focus of an attentional spotlight and are embedded in the discourse context.

The aim of the study was to conduct a qualitative sequential analysis of the temporal and semantic relations between clusters of the visual and the verbal data. To be able to do that, I created a temporally ordered multimodal score sheets (for details see Holsanova 2001:99f.). The score sheet in Figure 3 shows the visual behaviour (fixated objects as boxes on line 2) and verbal behaviour (verbal idea units on line 3), synchronised over time. Simple bars mark the borders of verbal foci (expressing the conscious focus of attention), double bars mark the borders of verbal superfoci (thematic clusters of foci that form more complex units of thought). This score sheet makes it possible to analyse what is happening during preceding, simultaneous and following fixations when a larger idea is developed and formulated.

With the help of this new analytic format, I could examine what was lying in the visual and verbal attentional spotlight at a particular moment: Configurations of verbal and visual clusters could be extracted and contents in the focused verbal idea flow and the visual fixation clusters could be compared. So for example, instead of analysing the *result* of the viewing of 'three birds in the tree' in form of

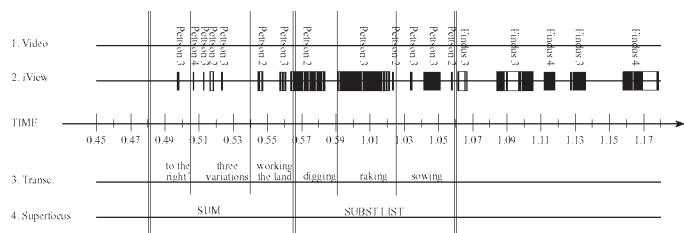


Figure 3. Multimodal time-coded score sheet. From Holsanova (2001:100).

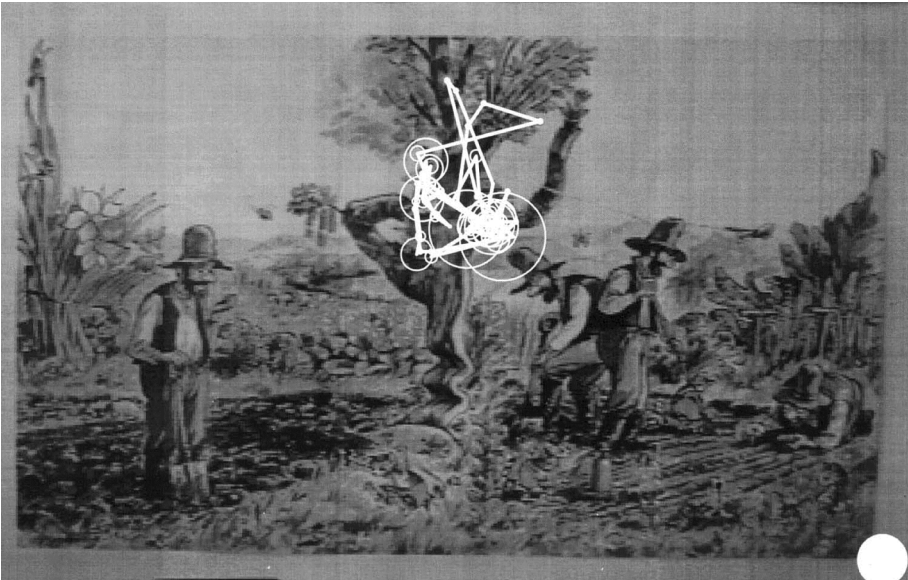


Figure 4. Fixation pattern: ‘Three birds in the tree’.

a fixation pattern (Figure 4) and the *result* of the picture description in form of a transcribed extract (see example 1), I was able to visualise and analyse the *process* of picture viewing and picture description on a time-coded score sheet (as seen in Figure 5). The multimodal method was applied in a recent project on perception and production in on-line writing (Andersson et al. 2006).

The fixation pattern shows us the path of picture discovery, that is, what objects and areas were fixated by the viewer and for how long. This is, though, a static pattern since it does not exactly visualise when and in what order they were fixated. The circles indicate only the position and duration of the fixations, the diameter of each fixation being proportional to its duration. The lines connecting fixations represent saccades. The white circle in the lower right corner is a reference point: it represents the size of a one-second fixation. Let us now look at the transcript.

Example 1. Extract from the transcript, translated from Swedish: ‘Three birds in the tree’:

0310	(2s) and in the middle of the field there is a tree	SUMMARIZING SUPERFOCUS
0311	. with one (1s) with . three birds’	
0312	. that are doing different things,	
0313	(1s) one bird is sitting on its eggs ’ in a nest ’	SUBSTANTIVE LIST OF ITEMS
0314	(1s) and the other bird (LAUGHING) is singing ’	
0315	at the same time as the . third . like female bird ’	
0316	(1s) is beating the rug or something,	

The transcript of the spoken language descriptions is detailed. It includes verbal features, prosodic features (such as intonation, rhythm, tempo, pauses, stress, voice quality, loudness), and non-verbal features (such as laughter). It also contains hesitations, interruptions, restarts false starts and other features that are typical of speech and that give us additional information about the speaker and the situational context. Each numbered line represents a new verbal focus expressing the content of active consciousness. Verbal focus is usually a phrase or a short clause, delimited by prosodic and acoustic features: it has one primary accent, a coherent intonation contour, and is usually preceded by a pause or hesitation (Holsanova 2001: 15f.). It implies that “one new idea is formulated at a time” (Chafe 1994: 108f.) and “active information is replaced by other, partially different information at approximately two second intervals” (Chafe 1987: 22). Several verbal foci are clustered into superfoci (for example summarizing superfocus or a list of items in the above example, delimited by lines). Verbal superfocus is a coherent chunk of speech, typically a longer sentence, that consists of several foci connected by the same thematic aspect and has a sentence final prosodic pattern. Superfoci can be conceived of as thresholds into a new complex units of thought.

As mentioned previously, in order to follow the dynamics of picture viewing and picture description and to analyse the temporal and semantic relations in this process, I had to synchronise the two data outcomes in time. Let us finally have a look at the new format in Figure 5.

In Figure 5, we are ‘zooming in’ on what is happening in the visual and in the verbal stream: which objects or areas are fixated visually and which areas are focused on verbally. As a result of the analysis of the temporal and semantic relations, I could extract several schematic configurations or patterns – both within a focus and within a superfocus (see Holsanova 2001: 102f.). Some of them can be seen above: for example an n-to-one configuration between the visual and the verbal part when the viewer is introducing the tree and the birds, or a series of delays between the visual and the verbal part, or a one-to-n configuration when the viewer is, later on, involved in a detailed description of the three different birds.

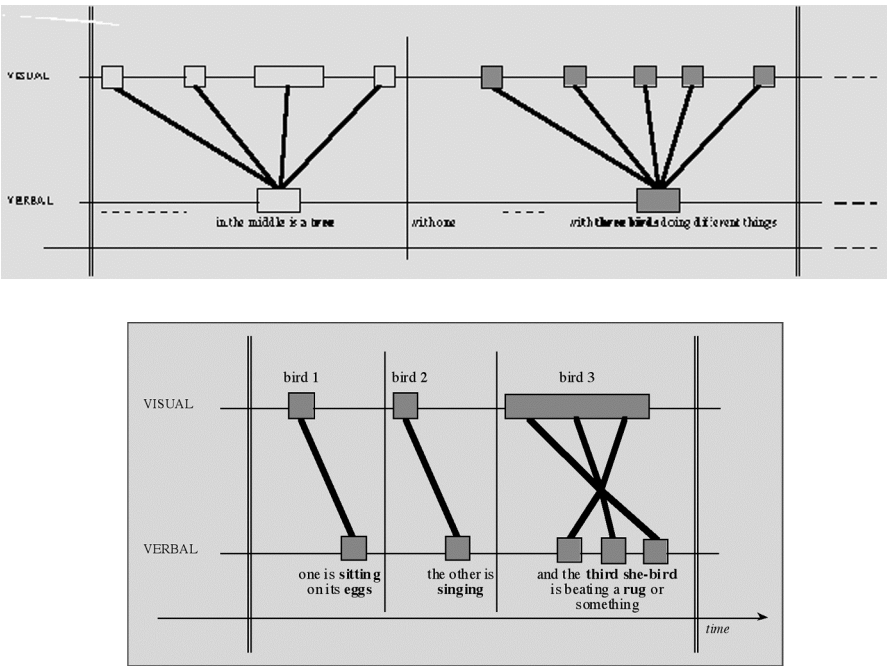


Figure 5. Schematic configuration on the multimodal time-coded score sheet: ‘Three birds in the tree’.

This section gave an introduction to the method and analytic tools used in a series of empirical studies. It was claimed that by combining spoken language description and eye movement protocols, we can get insights into what the observers found interesting, what drew their attention, and how the scene was perceived. The analytical tool – the multimodal time coded score sheet – enables us to synchronize visual and verbal behavior over time, to follow and compare the content of the attentional spotlight and to extract clusters in the visual and verbal flow. It is a processual visualization that gives us insight into the dynamics of picture viewing and picture description. The method developed is important for revealing principles involved in information processing (see Section 2). In addition, this method allows us to follow the process of bridging the gap between visual information gathering and speech production. In other words, we can examine, how the observers successively verbalize their thoughts about the picture (Section 3). Finally, the combination of the data method illuminates mental processes and attitudes and can thus be used as a sensitive evaluative tool for understanding of the dynamics of the ongoing perception process (Section 4).

Information processing: Temporal and semantic correlations between visual and verbal patterns

By now, the reader is acquainted with the method and the analytic tools, and we can review the results of the analysis. What we will be looking at in the following is the relation between the content of the visual focus of attention (specifically clusters of visual fixations) and the content of the verbal focus of attention (specifically clusters of verbal foci).

Three groups of questions are central: Can we identify comparable units in visual perception and in discourse production? Does the order of units in the verbal description reflect the general order in which information was acquired visually? Is the content of the units in picture viewing and picture description similar?

I was inspired by Wallace Chafe who claims that "... similar principles are involved in the way information is acquired from the environment (for example, through eye movements), in the way it is scanned by consciousness during recall, and in the way it is verbalised. All three processes may be guided by a single executive mechanism which determines what is focused on, for how long, and in what sequence" (Chafe 1980: 16). His thesis implies temporal and semantic correlation between the verbal and the visual data stream. In order to test it, I compared patterns during the visual scanning of the picture and during the verbal description of it. The aim was to find out what is lying in the visual and verbal attentional spotlight at a particular moment.

Before summing up the results, let us have a look at Figure 6. showing the process of picture discovery during the verbal formulation "... I see a tree in the middle ... and four men working around it". The superfocus (between the double bars) consists of two foci. Note that some objects are partly fixated during pauses in the spoken description (TREE) and some objects described in the second focus are partly fixated under the preceding focus (PETTSON 1, PETTSON 3). The consequence of this intertwined structure is that (a) there is no constant temporal relation between the visual and the verbal foci because of the partial lack of the verbal signal, (b) there is a preparatory visual fixation on Pettson 1 and Pettson 3 during the 'tree' – sequence. If we, however, look at the superfocus as a whole, we find agreement.

This example illustrates several principles that become revealed in the analysis of temporal and semantic relations in the focused verbal and visual flow: Visual foci track thought which is then verbalised selectively. To sum up, the results show that there is seldom a 1:1 relation between verbal and visual foci and a perfect temporal and semantic match or overlap between the visual and the verbal stream is very rare. The visual focus is often ahead of speech production. If one visual fixation is connected to one verbal focus, the verbal account is usually lagging behind the visual fixations with 2–3 seconds and this delay-configuration is a part

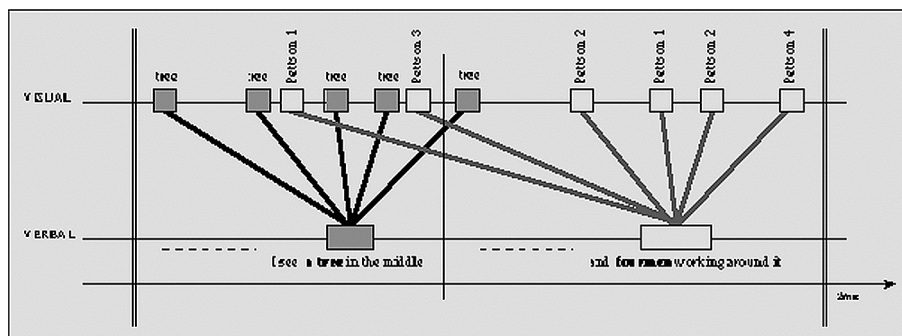


Figure 6. Complex temporal and semantic clusters: ‘Tree in the middle and four men working around it’.

of a larger unit (we saw the configuration ‘series of delays’ in Figure 4 where the three different birds were listed). Areas and objects are frequently re-examined which results either in multiple visual foci or in both multiple visual and verbal foci. Furthermore, the configurations where multiple visual foci are connected to one verbal focus are often intertwined and well integrated into larger wholes (as we could see in Figure 6). In addition, the results indicate that visual scanning is done both during pauses and simultaneously with the verbal description and thus, occasionally, there is no constant temporal correlation between the streams because of the partial lack of the verbal signal. In other words, if we want to find a clear correlation between spoken language and vision, we have to look at larger units of visual and verbal clusters. It seems that the wider the analytic discourse window, the larger the extent to which similar content can be observed in both data streams. I suggest that the superfocus is a suitable unit of comparison since it represents an entity that delimits separable clusters of both visual and verbal data.

In this section, I summarised and illustrated results of the temporal and semantic correlation between visual and verbal data. The theoretical approach used in the study was closely affiliated with functionalist approach and human information processing. The main result is that if we want to find temporal and semantic correlation between visual and verbal data, we have to look at higher levels of discourse hierarchy, that is, at the superfocus level.

We will come back to the semantic analysis in Section 4.

Variations in picture description: Two different styles

There are a number of interesting questions connected to picture viewing and picture description: how the viewers verbalize visual and spatial information from a non-linguistic source (mostly interesting for information-processing theories

and cognitive linguistics), how the observers bridge the gap between visual information gathering and speech production (mostly interesting for psycholinguists), how viewers structure and formulate their verbal description (mostly interesting for linguists and discourse analysts) and how viewers express their impressions from picture viewing, that is, how they describe their first-person experience from viewing (interesting for philosophers, phenomenologists and artistic sciences).

In this section, I will focus on the variation in picture descriptions since the results can be of interest for many of the above mentioned areas of research. In one of the studies where 12 informants described the picture from memory in an interactive setting for a listener), I found two different styles of description, deploying different perspectives. Attending to spatial relations in the picture was dominant in the static description style, while attending to the flow of time was the dominant pattern in the dynamic style. Consider the following two extracts:

- (a) 01 Well it's a **picture**'
 02 **rectangular**'
 03 and it's mainly **green** and **yellow**'
 04 with a **blue sky**'
 05 divided into **foreground** and **background**,
 06 in the middle there's a **tree**'
 uh and in it three **birds** are sitting,
 07 the lowest bird on the left sits on her **eggs**'
 08 and **above** her'
 09 there is a bigger bird **standing up**,
- (b) 01 It's quite early in the **spring**'
 02 and Pettson and his cat Findus are going to **sow**,
 03 Pettson starts with digging up his little **garden**'
 04 then he **rakes**'
 05 and . then . he sows **plants**'
 06 uh . puts **potatoes** out later on'
 07 and when he's **ready**'
 08 he starts sowing **lettuce**,

Extract (a) is a prototypical example of a static description style. In what could be called a technical description, the picture is typically decomposed into fields that are then described systematically, using a variety of terms for spatial relations. In the course of description, informants establish an elaborate set of referential frames that are used for localisations. They give a precise number of picture elements, stating their colour, geometric form and position. Apart from spatial expressions, the typical features of the technical description style are frequent use of nouns, existential constructions (*'there is', 'it is', 'it was'*), auxiliary verbs and passive voice.

Table 1.

Dynamic description style	Static description style
temporal perception	spatial perception
use of discourse markers	use of localisations
narrative schema	picture subdivided into squares described systematically
temporal adverbs, temporal subordinated clauses, temporal differences	numerous and precise localisations
dynamic motion verbs	nouns, position verbs (there is, there are)
dynamic, sequential description	static description

In the dynamic style (as in Extract (b)), observers primarily focus on temporal relations and dynamic events in the picture. Although there is no temporal or causal order inherent in the picture, viewers infer it: They explicitly mark that they are talking about steps in a process, about successive phases, about a certain order. The dynamic quality of this style is achieved by a frequent use of temporal verbs, temporal adverbs and motion verbs in active voice. Discourse markers are often used to focus and refocus on the picture elements, and to interconnect them. Apart from the above mentioned features, the informants seem to follow a narrative schema: the descriptions start with an introduction of the main characters, their involvement in various activities, and a description of the scene. Table 1 summarizes the characteristics of the static versus dynamic description style.

We are aware of the fact that non-linguistic aspects such as the informants' knowledge, interests and experience, may to some extent have influenced the description. Especially the informants' previous knowledge of the picture, of the story, and of the characters may be the most critical factor for distinction of the two description styles since not knowing the characters may lead to a less dynamic description. But when looking closely at the descriptions, we could exclude this factor since 10 out of 12 informants mentioned the characters by name.

Even the informants' expertise on painting techniques, farming, fauna and flora could have contributed to the way the picture was perceived and described. In addition, the informants' associations, their way of remembering things may have played role for the description. For this reason, the instruction and the interactional setting were kept constant and the background and the leisure-time interests have been accounted for in a questionnaire.

This section has been devoted to variations in the picture descriptions. Two different tendencies have been observed: the preference to focus on spatial relations in the picture typical of the static technical description style and the preference to focus on dynamic events and temporal relations in the picture typical of the dynamic narrative style.

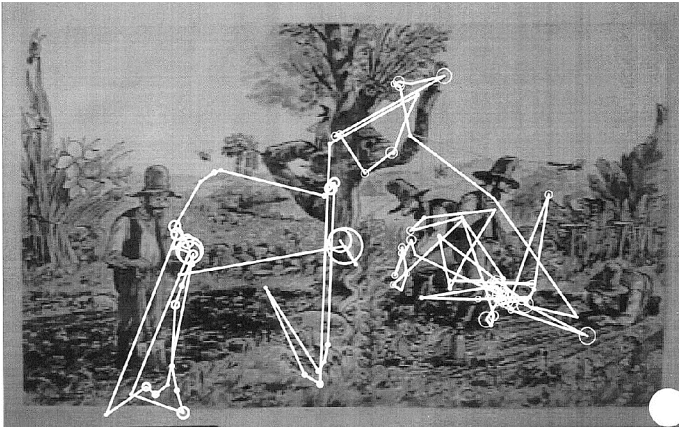


Figure 7. Fixation pattern for the first minute of picture viewing, informant 3.

Understanding the dynamics of the ongoing perception process

Unlike most of the studies in eye tracking and language production, associated with object identification and labelling on word or sentence level, the study of picture viewing and picture description focuses on higher levels of discourse. The starting point are complex ideas about the picture discovery that are formulated in large units of discourse. Because of this character, the verbal and visual data give us insights into the ongoing perception process. Apart from reporting about WHAT they see, the viewers also focus on HOW the picture appears to them. In other words, the informants are involved in both categorizing and interpreting activities.

In the scan path in Figure 7, we can see the way the picture was discovered and examined during the first minute by the one of the viewers.

In the transcript below we can the beginning of the verbal description performed by the same informant:

Transcript

0302	I see three/ eh four persons'
0303	who . are digging in a (2s) in . a field '
0304	or . one person is digging '
0305	(1s) another person is raking '
0306	(1s) a third person . is standing and looking at something
0307	at the same time as he is holding a spade '
0308	and a fourth person seems to sow things ,
0309	(2s) in this field,

0310	(2s) and in the middle of the field there is a tree
0311	. with one (1s) with . three birds'
0312	. that are doing different things,
0313	(1s) one bird is sitting on its eggs ' in a nest '
0314	(1s) and the other bird (LAUGHING) is singing '
0315	not because it is/
0316	at the same time as the . third . like female bird '
0317	(1s) is beating the rug or something,
0318	(1s) when I think about it
0319	then it seems as if it in fact were two different fields,
0320	(1s) one can interpret it as if they were in two different fields'
0321	those persons here,

This piece of description illustrates most of the activities that informants do when describing their impressions from the picture:

- They categorize the referents, states and events associated with the picture content;
- They evaluate the picture elements from a qualitative point of view (describe their colours, sizes and attributes);
- They specify the relations between the picture elements, compare picture elements and use metaphors;
- They consider the compositional aspects and group the picture elements in a certain way;
- They express their attitudes, associations, experience and report about mental states.

As for categorization, the main figure, Pettson, can be described as a person, a man, an old guy, a farmer, or he can be called by his name. His appearance could be described – *a weird guy* –, his cloths – *wearing a hat* –, the activity he is involved in can be specified – *he is digging* – and this activity can be evaluated: *he is digging frenetically*. Thus, we can envision a description on different levels of specificity and with different degree of creativity and freedom.

The tendencies that could be extracted from the data, are the following: Informants start either with a specific categorisation which was then verbally modified (*in a potato field or something; the second bird is screaming or something like that*) or, more often, a vague categorisation 'filler' is followed by a specification: *Pettson has found something he is looking at' he is looking at the soil*'. In other words, a general characteristic of an object is successively exchanged against more specified guesses: *'he is standing and looking at something' maybe a stone he has in his hand that he has dug out*'. When extracting an activity, for example when formulating *'Pettson is digging'*, the eye movements depict the relationships among picture elements

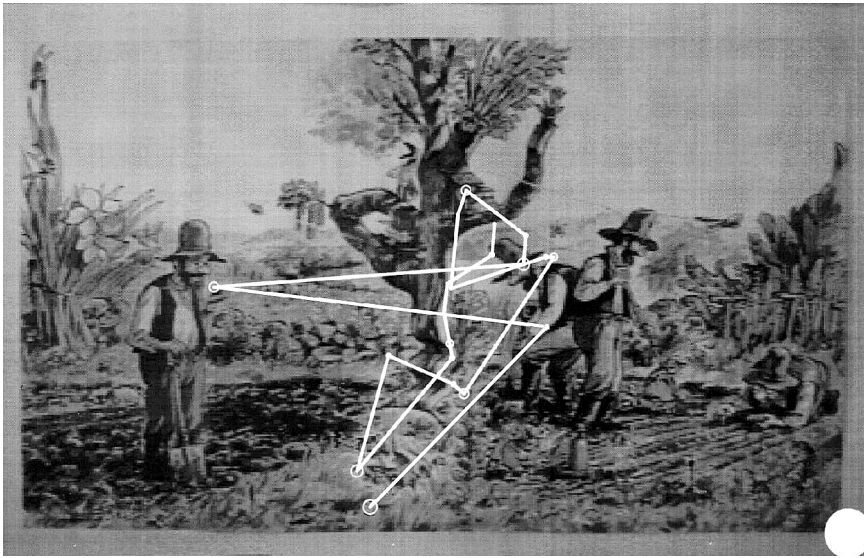


Figure 8. Fixation pattern: 'Two different fields'.

and mimic it in repetitive rhythmic units, by filling in links between functionally close parts of an objects (face – hand – tool; hand – tool, hand – tool – soil). In terms of cognitive semantics we can say that eye movements fill in the relationship according to a schema for an action.

During evaluations, the observers check the details (form, colour, contours, size) but also match the concrete, extracted features with the expected features of an prototypical/similar object: *'something that looks like a telephone line or telephone poles that is far too little in relation to the humans'*. They compare objects both inside the picture world and outside of it. By using metaphors from other domains they compare the animal world with the human one: *'one bird is looking very human'*; *'there's a dragonfly like a double aeroplane'*.

Concerning instrospection and report on mental states, let us look at the following example where one informant, after one minute of picture viewing and picture description, verbalises the following idea: *when I think about it then it seems as if it in fact were two different fields, one can interpret it as if they were in two different fields' those persons here.*

When inspecting the visual cluster (Fig. 8), we can see that the observer is rescanning several earlier identified objects that are distributed in the scene. This is consistent with Yarbus' (1967) finding that eye movements occur in cycles and that observers return to the same picture elements several times. This phenomenon has been confirmed by Noton and Stark (1971) who coined the word 'scanpath' to describe the repetitive fixations in scene viewing. What we do not see and know,

however, is how the observer perceives the objects on different occasions. I would claim that the objects that are refixated represent a bigger (compositional) portion of the picture and give support for his reconceptualization. By mentally zooming out, he discovers an inferential boundary between parts of the picture which he has not perceived before. The scene originally perceived in terms of 'one field' has become 'two fields' as the observer gets more and more acquainted with the picture. This example illustrates the dynamics of that the observer's perception of the picture unfolds dynamically and can change over time.

In sum, the combination of visual and verbal data shows that the objects are focused on and conceptualised on different levels of specificity, objects' location and attributes are described and evaluated, judgements about properties and relations between picture elements are formulated, metaphors are used as a means of comparison and finally, the object's activity is described. All this involves interpretation and creativity. What we have witnessed here is a process of stepwise specification, evaluation, interpretation and even reconceptualisation of picture elements and the picture as a whole.

Spatial, semantic and mental grouping

A second point I would like to make in this section is about grouping. According to Gestalt psychology, the organizing principles which enable us to perceive the patterns of stimuli as meaningful wholes are defined as (a) proximity, (b) similarity, (c) closure, (d) continuation, and (e) symmetry. The *proximity* principle implies that objects placed close to each other appear as groups rather than a random cluster. The *similarity* principle means that there is a tendency for elements of the same shape or colour to be seen as belonging together. Finally, the *symmetry* principle means that regions bounded by symmetrical borders tend to be perceived as coherent figures.

We can think of several principles of grouping that could have guided the infomants' picture description: spatial proximity principle (objects that are depicted closely to each other are described together), categorical or taxonomical proximity (elements that are similar to each other are described together), compositional principle (units delimited by composition are described together), saliency principle (expected, preferred and important elements are described first), animacy principle (human beings and other animate elements are described first) or even other principles.

Let us start with spatial proximity. When we consider a group of objects in the scene (for example the three birds in the tree), it is claimed that observers tend to fixate the closest objects first. This phenomenon is very coercive (because of the proximity to the fovea) and is described as an important principle in the

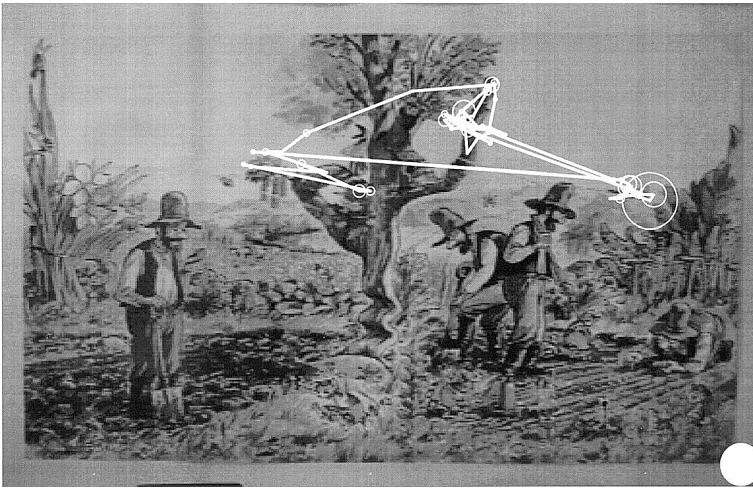


Figure 9. Fixation patterns for 'Flying objects'.

literature (Lévy-Schoen 1969, 1974, referred by Findlay 1999). And in fact, informants in my study seemed to be partly guided by this principle: they fixated clusters of objects depicted with spatial proximity. However, observers also fixated and described clusters which were not spatially close.

One type of cluster was based on grouping of multiple similar concrete objects (*four cats, four versions of Pettson*). This type of cluster could not be based on spatial proximity because the picture elements are distributed over the whole scene. In this case, we could rather speak about categorial proximity. (The simultaneity of the objects involved in different activities has probably promoted this guidance.) This type of cluster is usually described in a summarising superfocus that is typically followed by a list where each instance is described in detail.

Another type of cluster that was perceived as a meaningful unit in the scene was *hills on the horizon*. The observers' eyes followed the horizontal line, filling in links between objects. This cluster was probably a compositionally guided cluster. The observer is zooming out, scanning picture elements on a compositional level.

We are now moving further away from the scene inherent spatial proximity and coming to the next type of clusters constructed by the observers. This time, the cluster is an example of an active mental grouping of concrete objects based on an extraction of similar traits and activities (see Figure 9: *flying objects*). Despite the fact that the objects are distributed across the whole scene, they can be perceived of as a unit because of the identified common denominator. The observer is mentally zooming out and creating a unit relatively 'independent' of the 'suggested' meaningful units in the scene. The eye movements mimic the describers' functional grouping of objects.

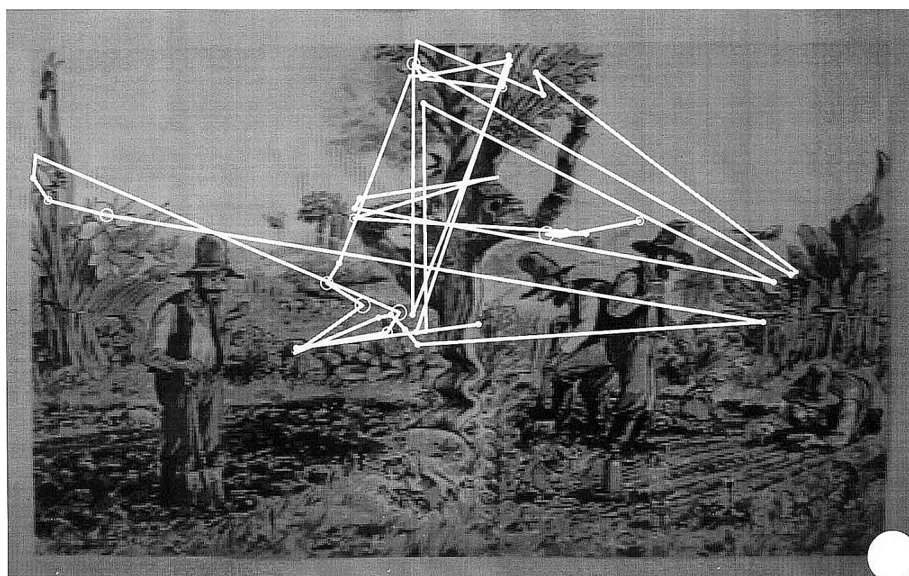


Figure 10. Fixation patterns by a person saying “It looks like in early summer”.

In a number of cases, especially in later parts of the observation, the clusters were based on thematic aspects and guided by top-down factors. In the following example (Figure 10), the observer is verbalising his impression about the picture content: “*it looks like an early summer*”. Previous scanning of the scene has led the observer to an indirect conclusion about the season of the year. In the visual fixation pattern, we can see large saccades across the whole picture composition. It is obviously a top-down guided cluster, based on mental coupling of concrete objects, their parts or attributes (such as flowers, leaves, colours) on a higher level of abstraction.

Since the concept “early summer” is not directly identical with one or several concrete objects in the scene, the semantic relation cannot be a mere referential one. Instead, the relation is inferred, the objects being concrete indices of a complex (abstract) concept. In addition, the relation between the spoken description and the visual depiction is not a categorical one, associated with object identification. Instead, the observer is formulating how the picture appears to him, on an abstract level. Afterwards, during visual rescanning, the observer is searching again for concrete objects and their parts as crucial indicators for his abstract statement. By refocusing these elements, the observer is in a way collecting evidence for his statement. In other words, he is checking whether the object characteristics in the concrete scene match with the symptoms for the described scenario. Concrete objects can be viewed differently on different occasions as a result of our mental zooming in and out. We have the ability to look at a single concrete object and

simultaneously zoom out and speak about an abstract concept or about the picture as a whole. When talking about creativity and freedom, this type of 'mental groupings' shows a high degree of creativity.

In sum, apart from scene inherent concrete picture elements with spatial proximity, even new creative groupings were described on the way. Objects across the scene – horizontally or vertically aligned – were grouped due to the scene composition. Multiple similar elements distributed in the scene were clustered on the basis of a common taxonomic category. Active mental groupings were created on the basis of similar traits and common activity. The process of mental zooming in and out could be documented, where concrete objects were refixated and viewed on another specificity level or with another concept in mind. During their successive picture discovery, the informants also created new 'mental' groupings across the whole scene, based on abstract concepts.

Picture viewing and mental imagery

In this section, we will be considered with mental imagery. Arnheim (1969), together with many others, puts the question whether we think in images and what form mental images possibly have. Finke (1989:2) defines mental imagery as "the mental invention or recreation of an experience that in at least some respects resembles the experience of actually perceiving an object or an event, either in conjunction with, or in the absence of, direct sensory stimulation." It has been proposed that we use mental imagery when we mentally invent or recreate personal experience, when we read novels, plan future events, retrieve information about the physical properties and relations of objects, imagine transformations by mental rotation and mental animation and when we solve problems. A large number of eye tracking studies have been conducted to test mental imagery.

Brandt and Stark (1997) and Laeng and Teodorescu (2001) have shown that spontaneous eye movements closely reflect the content and spatial relations from the original picture or scene. Holsanova, Hedberg, and Nilsson (1998) found similar tendencies as Brandt and Stark (1997), with the difference that the original picture showed a natural, real life scene. In order to extend the finding of previous studies, we chose Sven Nordqvist's (1990) picture of high complexity as our visual stimulus and conducted a number of new eye tracking studies (Johansson, Holsanova, & Holmqvist 2005, 2006). In the following, I will mention one of our experiments. Twelve students at the University of Lund, 6 females and 6 males, volunteered to participate in the experiment. They were asked to keep their eyes open so that we could film their pupil. The experiment consisted of two main phases, a *viewing* phase in which the subjects inspected the stimulus picture and a *description* phase in which the participants with their own words described this picture

from memory while looking at a white screen. Eye movements were recorded during both phases. At the beginning of the viewing phase, each subject received the following instructions:

“You will see a picture. We want you to study the picture as thoroughly as possible. While you study the picture we will measure your pupil size”.

The picture was shown for about 30 seconds, and was then covered by a white screen. The following description phase took usually 1–2 minutes. The subjects’ descriptions were transcribed in order to analyze *when* certain picture elements were mentioned. The eye movements were then analyzed according to objects derived from the descriptions. For instance, when a subject formulated a following superfocus,

- 01:20 – And ehhe to the **left** in the picture’
- 01:23 – there are large **daffodils**,
- 01:26 – it looks like there also sat some **animals there** perhaps,

we would expect the subject to move her eyes towards the left part of the white screen during the first focus. Then it would be plausible to inspect the referent of the second focus (the daffodil). Finally, we could expect the subject to dwell for some time within the daffodil area – on the white screen – searching for the animals (three birds, in fact) that were sitting there on the stimulus picture.

Eye movements were considered correct in *local correspondence* when within a certain time interval moving from one position to another in the correct direction. Eye movements were considered correct in *global correspondence* when moving from one position to another and finishing in a position that is spatially correct relatively the whole eye-tracking pattern of the subject (for a detailed description of our method, cf. Johansson et al. resubmitted). We used the Wilcoxon Signed-Ranks test for significance between the number of correct eye movements and the expected number of correct movements by chance.

As Table 2 shows, our results are significant both in the local and in the global correspondence coding suggesting that subjects visualize the spatial configuration of the scene as a support for their descriptions from memory. The effect we measured is strong. More than half of all picture elements mentioned had correct eye movements, according to the conservative global correspondence criteria. Allow-

Table 2. Picture description.

	% objects with correct eye movements	Direction significance
Local correspondence coding	74.8%	W = 76, z = 2.96 p = 0.0015
Global correspondence coding	54.9%	W = 66, z = 2.57 p = 0.0051

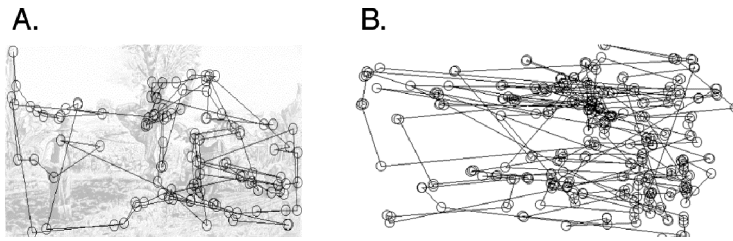


Figure 11. Same subject: Viewing phase (A) and description phase (B).

ing for re-centering and resizing of the image – as with local correspondence – makes almost three quarters of all picture elements have correct eye movements. Our data indicate that even for a complex picture, spatial locations are to a high degree preserved when describing it from memory.

The resizing effects, i.e. the fact that subjects shrunk, enlarged and stretched the image, were quite common during picture description. It was also common that subjects re-centered the image from time to time; thus yielding local correspondence. Overall, there was a good similarity between data from the viewing and the description phases, as can be seen in Figure 11.

The subjects' spatial pattern of eye movements was highly consistent with the original spatial arrangement. According to Kosslyn (1994), distance, location and orientation of the internal image can be represented in the visual buffer and it is possible to *shift attention* to certain parts or aspects of it. Laeng and Teodorescu (2001) interpret their results as a confirmation that eye movements play a functional role during image generation. Mast and Kosslyn (2002) propose, in a similar way as Hebb (1968), that eye movements are stored as spatial indexes that are used to arrange the parts of the image correctly. Our results can be interpreted as further evidence that eye movements play a functional role in visual mental imagery and that eye movements indeed are stored as spatial indexes that are used to arrange the different parts correctly when a mental image is generated.

There are, however, alternative interpretations. Researchers within the 'embodied' view claim that instead of relying on an internal image, we use features in the external environment. An imagined scene can then be projected over those external features and any storing of the whole scene internally would thus be unnecessary. Ballard et al. (1997) suggests that subjects leave behind 'deictic pointers' to locations of the scene in the environment, which later may be perceptually accessed when they are needed. Pylyshyn (2001) has developed a somewhat similar approach to support propositional representations and speaks about 'visual indexes'. Another alternative account is the 'perceptual activity theory' suggesting that instead of storing images, we store a continually updated and refined set of procedures or schemas that *specify how to direct our attention* in different situa-

tions (Thomas 1999). In this view, a perceptual experience consists of an ongoing, schema-guided perceptual exploration of the environment. Imagery is then the *re-enactment* of the specific exploratory perceptual behavior that would be appropriate for exploring the imagined object as if it were actually present. A somewhat similar approach is favored by Barsalou (1999).

In this section, I introduced a study in picture viewing and mental imagery. A significant similarity was found between (a) the eye movement patterns during picture viewing and (b) those produced during picture description (when the subjects were looking at a white screen). The eye movements closely reflected the content and the spatial relations of the original picture suggesting that the subjects created some sort of mental image as an aid for their descriptions from memory.

Summary and discussion

Current theories in visual perception stress the cognitive basis of art and scene perception. We “think” art as much as we “see” art (Solso 1994). Our way of perceiving objects in a scene can be triggered by our expectations, our interests, intentions, previous knowledge, context or instructions.

The aim of the presented studies was to show how overt verbal and visual protocols can, in concert, elucidate covert mental processes. This combination of data illuminates mental processes and attitudes and can thus be used as a sensitive tool for understanding of the dynamics of the ongoing perception process. The results can in this way also contribute to the area on the intersection of artistic and cognitive theories. My data suggest that it is not only recognition of objects that matters but also how the picture appears to the viewers. Verbal descriptions include quality of experience, subjective content and mental state. Viewers report about (i) referents, states and events, (ii) colours, sizes and attributes, (iii) compositional aspects, (iv) they mentally group the perceived objects into more abstract entities, (v) compare picture elements, (vi) express attitudes, associations, (vii) report about mental states.

Subjects start by looking at picture-inherent objects, units and gestalts. As their picture viewing progresses, they tend to create mental units that are more independent of the concrete picture elements. They may make large saccades across the whole picture, picking up information from different locations to support concepts which are distributed across the picture (like “early summer” or “flying insects”). With the increasing cognitive involvement, observers and describers tend to return to certain areas, change their perspective and reformulate or recategorise the scene. It becomes clear that perception of the picture changes over time. The dynamics of this categorization process is reflected in the usage of many re-fixations in picture viewing and reformulations, paraphrases and modifications in

picture description. Finally, during description from memory, the subjects seem to “redraw” or “reconstruct” certain features from the original picture with the help of mental imagery.

References

- Andersson, B., Dahl, J., Holmqvist, K., Holsanova, J., Johansson, V., Karlsson, H., Strömquist, S., Tufvesson, S., & Wengelin, Å. (2006). Combining keystroke logging with eye tracking. In Luuk Van Waes, Marielle Leiten & Chris Neuwirth (Eds.), *Writing and Digital Media*. Elsevier BV (North Holland).
- Arnheim, R. (1969). *Visual Thinking*. Berkeley and Los Angeles: University of California Press.
- Arnheim, R. (1974). *Art and Visual Perception. A Psychology of the Creative Eye*. Berkeley, Los Angeles, London: University of California Press.
- Ballard, H. D., Hayhoe, M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral Brain Sciences*, 20, 723–767.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.
- Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience*, 9, 27–38.
- Buswell, G. T. (1935). *How People Look at Pictures. A Study of the Psychology of Perception in Art*. Chicago: The University of Chicago Press.
- Chafe, W. L. (1980). The deployment of consciousness in the production of a narrative. In W. L. Chafe (Ed.), *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production* (pp. 9–50). Norwood, NJ: Ablex.
- Chafe, W. L. (1994). *Discourse, Consciousness, and Time. The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago, London: The University of Chicago Press.
- Chafe, W. L. (in press). Language and consciousness. *Encyclopedia of Consciousness*. Oxford.
- Finke, R. A. (1989). *Principles of Mental Imagery*. Cambridge, MA: MIT Press.
- Hebb, D. O. (1968). Concerning imagery. *Psychological Review*, 75, 466–477.
- Holsanova, J. (2001). Picture viewing and picture description: Two windows on the mind. Doctoral dissertation. *Lund University Cognitive Studies*, 83.
- Holsanova, J., Hedberg, B., & Nilsson, N. (1998). Visual and verbal focus patterns when describing pictures. In Becker, Deubel, & Mergner (Eds.), *Current Oculomotor Research: Physiological and Psychological Aspects*. New York, London, Moscow: Plenum.
- Johansson, R., Holsanova, J., & Holmqvist, K. (2005). What do eye movements reveal about mental imagery? Evidence from visual and verbal elicitations. *Proceedings of the Cognitive Science Conference*. Stresa, Italy.
- Johansson, R., Holsanova, J., & Holmqvist, K. (2006). Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness. *Cognitive Science*, 30 (6). Lawrence Erlbaum.
- Kosslyn, S. M. (1994). *Image and Brain*. Cambridge, MA: The MIT Press.
- Laeng, B., & Teodorescu, D.-S. (2002). Eye scanpaths during visual imagery reenact those of perception of the same visual scene. *Cognitive Science*, 26, 207–231.
- Mast, F. W., & Kosslyn, S. M. (2002). Eye movements during visual mental imagery. *TRENDS in Cognitive Science*, 6 (7).
- Nordqvist, Sven (1990). *Kackel i trädgårdslandet*. Opal.

- Noton, D., & Stark, L. W. (1971a). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, 11, 929–942.
- Noton, D., & Stark, L. W. (1971b). Scanpaths in eye movements during perception. *Science*, 171, 308–311.
- Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, 80 (1/2), 127–158.
- Solso, R. L. (1994). *Cognition and Visual Arts*. A Bradford Book. Cambridge, MA, London: England, The MIT Press.
- Thomas, N. J. T. (1999). Are theories of imagery theories of imagination? An active perception approach to conscious mental content. *Cognitive Science*, 23 (2).
- Yarbus, A. L. (1967). *Eye Movements and Vision* (1st Russian edition 1965). New York: Plenum Press.

Order and complexity in naturalistic landscapes

On creation, depiction and perception of Japanese dry rock gardens

Gert J. van Tonder

Introduction

In a discussion of order and complexity of a design, the lack of a clear definition of the two concepts soon becomes apparent. It nevertheless proves useful in an attempt to grasp the visual psychology of gardens. What I intuit as order and as complexity relate closely to the *subjective experience* of what I am convinced is nothing short of masterpieces of visual depiction. I am talking about the act of looking at, or rather, looking *into* a classical Japanese dry rock garden (*karesansui*)¹ design.

When applied to dry rock gardens, formal definitions of order and complexity – especially quantitative measures – fail to confirm what I perceptually experience as compelling. No specific grid, ratio, clearly identifiable fractal dimension or other measure reveals why I distinctly experience the garden as a calming space innervated by a subtle balance of tensions. When a given measure deals with spacing of elements, it fails to account for asymmetries; or when accounting for that, it is insensitive to gradients or alternations, or symmetries and other visual logic that comes into play as the multitudinous interactions between different visual mechanisms reach equilibrium. Formal quantitative measures of order and complexity end short of where perception begins, and seem to exclude my perceptual experience of complexity and order in the garden. And yet I am looking for quantifiable measures of this experience, for I am convinced that it has a legitimate objective aspect worth exploring.

One could consider the various philosophical and cultural interpretations of gardens. Arrangements of rocks may symbolize written Chinese letters, felicitous symbols like turtles, cranes, boats and gourds, or geological features, such

as mountains or islands, evocative of legends or moral tales. The garden can be viewed as the space where seasonal changes in nature are accentuated: the flowers of spring; summer green seeping through every crack; autumn wind in the drying grass; monochrome winter silence. Valid as this approach is in its own right, it does not lead me clearly towards the visual psychology of the garden. Speaking only on my own behalf, the most profound insights follow from continued direct perception of the garden as it is, without deliberately seeking for the subtle semblances of dragons or tigers in the shapes of rocks.

Investigation of the relation between gardens as visual stimuli and activation of perceptual mechanisms is supported by the casual observation that the garden captivates even naïve garden visitors who may lack detailed knowledge of Japanese culture and philosophy. And it is not simply that they mistakenly project their own cultural expectations onto a novel design: some aspect of the garden as visual pattern seemingly finds rapport with the perceptual apparatus. There should thus be a suitable level for understanding the process of garden viewing as direct perception, where the garden is interpreted as the stimulus pattern on which the universal part of perceptual human (or primate?) brain acts. Universality in perception arises insofar all ecological realms of human visual perception share universal aspects. I hope to shed more light on how visual structure in *karesansui* is related to activation of the universal brain, and in particular, why it perceptually leads to a quality of non-trivial calm.

One cannot rule out the possibility that the abovementioned existing quantitative and philosophical approaches will prove to be more useful in future investigations looking at perception in terms of order and complexity in gardens. At present, three other types of sources informed the research presented here. They are (1) definitions of order and complexity derived from accomplished visual intuition, (2) classical Japanese gardening guidelines and (3) the corpus of perceptual grouping factors as studied within the Gestalt approach to vision.

Intuitive definitions of order and complexity

In his essay on order and complexity, Rudolph Arnheim (1966) defines order as the degree and kind of lawfulness governing the relations among the parts of an entity, while complexity is the multiplicity (described by Arnheim to mean not a mere duplicity, but multifariousness) of relations between parts. In his approach, order and complexity are antagonistic yet have no existence independent of each another. Complexity alone creates confusion; isolated order invites boredom; great art is rich in both order and complexity. Arnheim defines disorder as the structural condition that lacks the order necessary for optimal human functioning – sometimes needed to create a sense of freedom in design. Order in landscapes varies with the

degree of their *definition*, where the latter indicates the extent to which order is carried through in a design. He stresses four qualities of order, namely *homogeneity* (low level of structural complexity, resulting from different orders canceling one another), *coordination* (higher level of equal structural complexes, resulting from balanced orders), *hierarchy* (structural complexity arising from clearly related orders) and *accident* (the order of independence). While these qualities refer to static spatial arrangements of objects, Arnheim claims that structures come to life due to the dynamic of perception. While the above rings true to one's visual intuition, their implementation in a formalized analysis of visual structure is by no means clear. I find Arnheim's distinctions instructive in the search for a deeper understanding of perceptual effects associated with garden design and will return to his concepts again in a later section.

In a similar vein, Christopher Alexander creates a definition of order according to his collected observations of natural and man-made structures (Alexander 1979, 2002). He arrives at fifteen structural qualities distilled from patterns which he observed over a life time of architectural work and intuitively experience as *alive*. The qualities are intended to describe how (structural) centers reinforce each other, and how centers become incorporated into a whole. The qualities are (1) levels of scale, (2) strong centers, (3) boundaries, (4) alternating repetition, (5) positive space, (6) good shape, (7) local symmetries, (8) deep interlock and ambiguity, (9) contrast, (10) gradients, (11) roughness, (12) echoes, (13) the void, (14) simplicity and inner calm, and (15) not-separateness. Similar to Arnheim, Alexander gives many examples of what his qualities encompass, but these insights do not all translate into evident analytical measures. This is not to invalidate his observations, but rather to say that such measures still have to be uncovered. Granting Alexander the benefit of the doubt, the approach is thought provoking insofar as it provides a framework for understanding order.

Classical Japanese gardening principles

Dealing with order and complexity in a different manner, Japanese garden designers intended to create visual spaces that depict a canonical view of idealized nature. In this process the gardener pursues various aesthetic ideals, such as naturalness, simplicity, asymmetry and tranquility – relatable to visual structure – and more metaphysical descriptions of gardens, such as profundity, deep solitude and subtle confoundedness (Nitschke 2000). As with Arnheim and Alexander, gardening principles therefore contribute mainly metaphorical descriptions of the concepts of order and complexity in landscape design. The main corpus of classical design guidelines derives from the collected teachings of the 11th century monk *Zouen* (Shingen 1466) and the Heian period (794–1132 AD) garden design

manual, *Sakuteiki* (Shimoyama 1976). An example of a design guideline from a translation of *Sakuteiki* (Jiroh & Keane 2001) reads as follows:

If there are stones that “flee”, then there should be stones that “chase” after; if there are stones that lean, then there should be those that lend support; if some proceed then others should acquiesce; if some face up, then others should face down; and to balance with stones that stand upright there should also be those that recline.

Without explicitly knowing these manuscripts, any salted Japanese gardener today would be familiar with the classical design guidelines, as it forms the staple of concepts in the creation of a Japanese garden. Otherwise, various instructions exist as part of an oral tradition.

Chinese garden design developed after classical landscape painting and consequently shows many features of the depicted idealized landscape. Various waves of Chinese cultural influence shaped Japanese society during the medieval eras. Notably, Chinese painting and landscaping had a huge impact on Japanese arts. Indigenous Japanese religious and other practices gradually transformed the Chinese style to suit local conditions. The elaborate Chinese gardening nomenclature became much condensed, but without sacrificing the original link with depictive space of the landscape. In Japanese architecture nearly every space has an associated viewpoint, and appears framed, even if just by an alcove, window, sliding door or other device, and associated viewpoint(s).

In a nutshell, the classical guidelines state that the ultimate task of the gardener is to convey the essence of nature through careful observation of naturalistic patterns, and present the owner of the estate with felicitous surroundings in the form of a vision in which all elements are in good agreement. Rocks and trees constitute the backbone of the garden composition. As static and dynamic elements, respectively, each needs to be treated in accordance to a well worked out set of design effects to ensure that the whole captures the essence of nature while arising from parts which are in good agreement. Detailed description of the garden design guidelines should convince the reader of the close relation between garden creation and the concepts of order and complexity.

Perceptual grouping factors

The Gestalt school investigated the structural arrangements which evoke or dissolve perceptual ordering processes (Koffka 1935; Wertheimer 1938). Their framework brings us one step closer to a new understanding of visual effects in *kare-sansui*, serving as bridge to interpret classical design rules such as those which describe how parts are brought into good agreement with one another. Ultimately one would like to understand design effects not just in terms of structural levels, but also in terms of relations between depicted and perceived structure. Here

the Gestalt principles prove useful. Thus, figure-ground concepts such as boundary completion, filing-in of surface properties, completion of occluded or missing details and more naturally belong to this discussion.

The rest of the text is organized mainly into two parts. The first, consisting of four subsections, deals with the static component of the backbone of the garden, namely rocks. The dynamic components, i.e. trees and plants, are the topic of the second part, addressing perceptual grouping and techniques for shaping arboreal materials. The section summarizes the status quo of this investigation, which aims to formally interpret the effect of skilled versus unskilled pruning on visual structure of trees.

Rocks

The following four sections summarize our research on the visual structure of rock composition in *karesansui* gardens.

Classical gardening guidelines, Gestalt factors, depictive effects and Arnheim's notion of order and complexity

During training with the *Ueoto* gardening school in Kyoto between 1997–2000, the author became convinced of a rich link between intuitive visual effects applied in Japanese gardens and the principles used by the Gestalt school of psychology to predict the relationship between visual cues and human perceptual grouping. The first and second columns of Table 1 summarize the interpretations of that investigation (see Van Tonder & Lyons 2005 for details). We emphasized the similarities between visual effects observed both by the Gestaltists and classical Japanese garden designers who preceded the Gestalt school with about 400–900 years. Apparently Japanese gardeners aim towards subtle *balancing* of visual salience of design components at various spatial scales, whereas the Gestalt school aimed to *predict* perceptual grouping effects from different attributes of visual cues. *Karesansui* gardens can therefore be considered as sophisticated albeit intuitive perceptual stimulus patterns. Figure 1 (adapted from Van Tonder & Lyons 2005) illustrates how some garden design elements can be viewed within the Gestalt framework. The figure shows *Dokuzatei* (garden of solitary retreat) at *Zuihoin*, subtemple of *Daitokuji*, Kyoto (Figure 1A). This garden is a relatively modern construction by Shigemori Mirei, dating from 1961, yet created within the framework of classical design guidelines. Note the textural structure, visual junctions and silhouette shapes of rock surfaces, rocks, and rock clusters (Figure 1B1~B4). Here, Figure 1B4 shows a horizontally compressed image of the garden, to clarify global structure of rocks as backbone of the composition. For clarity, the textural patterns,

junctions and shapes are highlighted in Figure 1C1~C4. Examples where classical rules have not been followed demonstrate how symmetry, enclosure and good continuation, combined with high contrasts cause a shift in salience that compromises the sense of calm and visual harmony (Figure 1B5 and C5). The author takes it one step further (tongue in the cheek!) to violate the classical guidelines for naturalistic garden design in a not-so-calm version of the original garden (Figure 1D). The result is a nearly scintillating visual rivalry.

The most commonly used depictive effects in gardens are the rich layering of occlusion junctions, finely tuned or exaggerated size gradients, expansive homogeneous regions that function as regions of 'empty canvass', occlusion of the majority of the visual figure, and emphasis on occlusion junctions between figure and ground, seemingly to evoke the perceptual completion of occluded shapes. Subtle red-blue colour combinations and self-similarity between rock texture, rock silhouette and rock clusters adds a rich depictive reference of many possible spatial scales, especially that of large scale landscape scenery. Psychophysics show that similar conditions lower perceptual effort (Navon 1977; Kinchla & Wolfe 1979).

One finds that gardening design guidelines do not simply present a set of visual effects to enhance similarity or reduce bilateral symmetry. Nor does it only allow an increase of structural order or a decrease of complexity. As expected from a set of rules for subtle balancing of visual grouping effects, the design guidelines offer many pathways to either increase or reduce the effect of any visual attribute. All attributes and their consequences are interrelated and hence the need for this type of feedback. Note that most of the design guidelines affect more than one attribute simultaneously. In this lies one of the most difficult aspects of successful *karesansui* design.

The structure of MA

Arnheim visited Kyoto for about one year, during which time he was deeply inspired by Japanese architecture and landscape design. His visit to Japan culminated in the writing of an essay on order and complexity in naturalistic landscapes, where he considers the ultimate example of *karesansui* – the stone garden at *Ryoanji* in Kyoto. Arnheim concluded that the profound sense of order in this garden derives from the fact even if the rocks within each cluster form part of a small hierarchy, all the rock clusters are coordinated with one another without fitting into a global hierarchical structure. This brings *karesansui* in stark contrast to classical French gardens, for instance.

When considering the visual salience of empty space in *karesansui*, a different interpretation is possible. We have previously developed a computational model for image segmentation using forward-inverse medial axis transformation (Van

Table 1. Summarized comparison of classical Japanese design instructions on rock composition (column 1), Gestalt grouping factors (column 2), depictive effects (column 3) and Arnheim's concept of order and complexity (column 4) (Van Tonder 2004a–b; Van Tonder & Lyons 2005).

Design guideline	Gestalt factors	Depictive effect	Order and complexity
1 Regard rocks (and trees) as the backbone of a garden composition.	All	Garden construction requires a global design visualization.	Global order, global structural axis of garden. Global hierarchy.
2 The largest rock of each cluster is usually set first.	Proximity, breaking good continuation at large spatial scales.	Comparison of salience of global and local parts during construction.	Increase order, reduce global complexity, initialize hierarchy.
3 First set numerous large and small stones temporarily, with some upright and others reclining, showing the best side of each rock.	All	Design feedback through which local and global features can be consolidated.	Temporary inventory of potential orders and complexities as suggested by parts.
4 In the garden, each rock should be placed as it was found in nature.	Similarity.	A possible means to match the overall rock shape to its surface features.	Increase order and complexity. Clarification of hierarchy. Reduce clashing and uncoordinated orders.
5 Maximum rock size depends on the size of the garden courtyard.	Proximity, simplicity.	Setting of texture density such that textural crowding is avoided.	Increase order.
6 Various ratios of large:small rock size result in harmonious composition. For example, ratios of 3:1 and 3:2 are often used.	Proximity, similarity, asymmetry.	Enhance the degree of triangularity and asymmetry. Using roughly set ratios regularize the building blocks used throughout the design.	Increase order. Complexity increased by the use of diverse triangular cluster envelopes.
7 Spaces between rock clusters depend on size and height of each cluster. Also see #8. As a rule of thumb, the space between the centers of two clusters is roughly equal to twice their average size.	Proximity, similarity.	At this texture density, pop-out due to texture gradients is minimal (Van Tonder & Ejima 2000; Van Tonder & Lyons 2005).	Increase order, clarify hierarchy and coordination.
8 Larger rocks should be surrounded by accordingly larger open spaces, and the converse is true for smaller rocks.	Proximity, similarity.	Tessellation density of elements remains more or less equal throughout the garden.	Increase order, clarify hierarchy and coordination.
9 In a deep courtyard the entire rock composition should be placed closer to the viewer, to reduce the sense of spatial depth.	Proximity, simplicity.	Reducing spatial depth in the garden in order to lessen the extent of optical accommodation required to visually inspect each component.	Increase order.
10 In a courtyard with shallow visual depth, the rock clusters should be distributed throughout the courtyard, from front to back.	Proximity, simplicity.	Enhancing the sense of depth and spatial balance in the garden as visual scene.	Increase order.
11 Use components with distinct vertical, horizontal and diagonal lines.	Simplicity, similarity.	Simplify compositional units. Reinforce bilateral asymmetry. Clarify relations between rock shape, rock texture and rock cluster.	Reducing complexity, increasing order.

Table 1. (*continued*)

Design guideline	Gestalt factors	Depictive effect	Order and complexity
12 Use triangular rock shapes.	Simplicity, similarity over multiple spatial scales.	Triangles relate individual rocks to shapes of rock clusters at a larger spatial scale. Triangular rocks can be easily set to have a wide base at ground level, important for reasons explained in guideline #19.	Reduced range of orders. Clarified order.
13 Triads (clusters of three rocks) are preferable units of composition.	Simplicity, closure (triangles are simple enclosed figures), (Self) similarity.	Triangular clusters are the simplest closed figures in which grouping effects can be adjusted at any one time	Reducing the range of orders, i.e. reducing complexity. Clarifying remaining orders, clarifying hierarchy.
14 Using odd-numbered groupings of rocks; the total number of rock clusters should also be an odd number.	Breaking of good continuation, similarity over numerous spatial scales.	Creating trilateral visual junctions, i.e. strong occlusion junctions (monocular depth clues) and boundary ownership.	Increasing both order and complexity.
15 Clusters of 3, 5 and 7 rocks are often used.	Simplicity.	Ensure that clusters are varied but of manageable size.	Reducing complexity, increasing order.
16 All components should be bilaterally asymmetrical.	Breaking bilateral symmetry.	Increases compositional irregularity over various scales. Fine tune visual salience of individual objects, clusters and whole composition.	Reduce highly significant order.
17 Rocks with many angular facets are preferable.	Enhance self-similarity, reduce continuation and symmetry.	Ensure irregularity of rock silhouettes.	Increase order of hierarchy and coordination, increase complexity
18 Overall shape and textural grain of rocks should match.	Similarity.	Appropriate surface facets resemble occlusion junctions and entire rock shape (#14, #20, #23). Increased degree of self-similarity in the visual structure of the rock.	Clarify hierarchy. Reduce clashing orders. Increase complexity.
19 Plant rocks firmly, with two thirds of each rock buried underground.	Enhance good continuation and bilateral symmetry, weaken simplicity, breaking closure.	Outward sloping occlusion junctions between rocks and gravel give impression of a wider solid structure extending beneath the gravel.	Increase order i.e. global hierarchy. Reduce order at the level of the individual object.
20 Rocks depicting high mountains should overlap like the scales on a fish, or similar to the panels of a folding screen.	Breaking good continuation, enhancing multi-scale similarity and increasing simplicity.	Hierarchical occlusion junctions emphasize depth. Subtle variation of trilateral visual junctions adds to interest of visual texture.	Increase complexity at global level. Increase order. Reduce complexity. Increase hierarchy.
21 Rocks at the foot of a hill should appear like frolicking animals.	Breaking similarity, symmetry, good continuation and simplicity.	Reduce visual dominance of junctions in #20, this pattern visually bridges orderly visual junctions and noisy texture of the gravel.	Increase complexity, reduce order. Break hierarchy, increase coordination and accident.
22 At the foot of a hill (the lower part of a rock cluster) there should be rocks leaning against one another, in opposite directions.	Weakening good continuation, increasing bilateral asymmetry.	Similar to #21, the effect lends a structural transition from the largest rocks in the cluster to the surrounding vegetation and gravel.	Increase complexity, tuning hierarchy.
23 Landscape features, e.g. rivers, must never follow a straight path.	Breaking good continuation, closure and bilateral symmetry.	Increased irregularity lowers visual salience. Increased occlusion enhances perceived visual depth.	Reducing simplistic order, increasing complexity, increasing hierarchy and accident.

Table 1. (*continued*)

Design guideline	Gestalt factors	Depictive effect	Order and complexity
24 When the overall design is decided, smaller stones are added in good agreement with the larger stones.	Increase the degree of bilateral asymmetry, fine tune proximity and the degree of similarity within and between clusters.	Feedback during construction to equalize salience of design elements at larger and smaller spatial scales, and to improve shapes and visual junctions between larger components.	Increase complexity, while adjusting order at different scales. Increase hierarchy.
25 Any rock cluster should suggest ten thousand mountains in a single glance.	Self-similarity, careful tuning of proximity arrangement.	Unsaturated hues, homogenous textures, repetitive shapes, hierarchical occlusion junctions and empty visual ground emulate various depth effects, suggestive of expansive landscape scenery.	Clarification of hierarchy and coordination. Increased order and complexity.
26 Smaller rocks can be added to give the base of the main rock a more solid appearance.	Increase bilateral asymmetry, enhance good continuation, prevent closure.	Correcting the occlusion junctions between rocks and gravel as discussed in #19, #20, #21 and #22.	Increasing order and complexity at global levels.
27 <i>Suteishi</i> (low inconspicuous stones) give the composition a more natural appearance and draw rock clusters together.	Proximity, breaking bilateral symmetry, breaking of good continuation.	Visual unity of the entire design is fine tuned and visual salience of individual clusters adjusted.	Increase complexity, reduce simplistic order, increase accident.
28 Do not idealize your composition based on scenery from rocky shores, where boulders are constantly overturned.	Simplicity principle completely discarded.	Formations arising from turbulent natural forces do not convey the visual appearance suitable for calm harmonized garden design.	Various unrelated orders are mixed. Very high complexity.
29 A rock at the very back of a garden enhances the sense of extensiveness of the courtyard.	Proximity, increase of good continuation.	The rock composition may appear to extend beyond the garden courtyard.	Extend spatial range of orders and complexity within the garden composition.
30 Materials with striking textures should be avoided.	Simplicity, similarity.	Homogeneous textures reduce internal contours on object surfaces. Thus, a reduced number of visual parts raises salience of the whole.	Increase homogeneity, reduce clashing and uncoordinated orders. Increase complexity at smallest spatial scales.
31 <i>Shirakawazuna</i> (lit. White-river sand, a type of rough felspar gravel found in Kyoto) is the preferred gravel for kareansui.	Similarity, simplicity, weakens good continuation, breaks closure.	Bright, uniform texture increases salience of visual ground while adding irregular, fine grain texture to the design. This texture serves as unified occlusion surface between rock clusters.	Increase homogeneity, reduce unrelated or clashing order. Increase complexity (compared to plain white sand, which lowers complexity).
32 No part of the garden should be left untouched. Materials such as moss and gravel should fill even the last square centimeter.	Simplicity, good continuation.	Enhance the depictive quality of a garden through empty (unpainted) regions. Textures covering the entire garden are more dramatic and convincing in their effect as occlusion surfaces.	Reduce complexity. Increase order by reduction of unrelated orders.
33 Unsaturated colouration is preferred.	Simplicity, similarity.	Reducing visual salience of parts if compared to that of the whole.	Increase homogeneity, lower complexity.
34 Slight red and blue hues lend subtle impressions of spatial depth.	Break good continuation, tune perceived proximity.	Suggestive of atmospheric perspective, while visual salience of individual rocks is increased relative to that of their parent clusters.	Increases order and complexity at various spatial scales.
35 The composition should face the viewer from a slight angle, and not from directly ahead.	Break symmetry.	Enhance the sense of looking into the garden.	Increase complexity.



Figure 1. (A) Dokuzatei, Kyoto. (B1~B4) Texture, shapes, triangular frameworks and junctions between rocks (highlighted in C1~C4 for clarity). B5, C5 show examples where classical rules have not been followed. The author tampered with Dokuzatei in D to demonstrate the consequent disruption of visual calm.

Tonder & Ejima 2000). The method offers an interesting explanation to various perceptual phenomena linked to the image segmentation problem. Figure 2A~C shows three visual patterns and their corresponding segmentation based on forward-inverse medial axis transformation (unfortunately computed only at very low spatial resolution in this case). The method segments the images into meaningful parts, while simultaneously producing abstract shape descriptions in the form of medial axes. The means for computing medial axes in this model has been specifically adapted to replicate psychophysics related to medial axes (Kovács, Feher, & Julezs 2000) in a scale invariant, noise resistant manner (Van Tonder & Ejima 2003). Figure 2D1~D3 illustrates model operation in high noise resistant, intermediate and noise sensitive modes, respectively. Figure 2D2 most closely resembles the known psychophysics. All medial axis results discussed hereafter were produced with the same parameter settings in which Figure 2D2 was obtained.

When applying our adapted medial axis transformation to the empty space of *Ryoanji* (Figure 2E2) a clearly hierarchical structure becomes apparent (Van Tonder, Lyons, & Ejima 2002): A global branching structure with a strict branching rule, evenly distributed branching points and branches that all slope towards the viewer converges onto the canonical viewpoint of the garden. To show that the computation does not solely depend on contour strength (we initially set the contour values according to the height of each rock to roughly account for perceptual salience of larger rocks), the simulation is repeated for a map the garden where all contour pixels are set to a unitary value. The global arrangement of first order

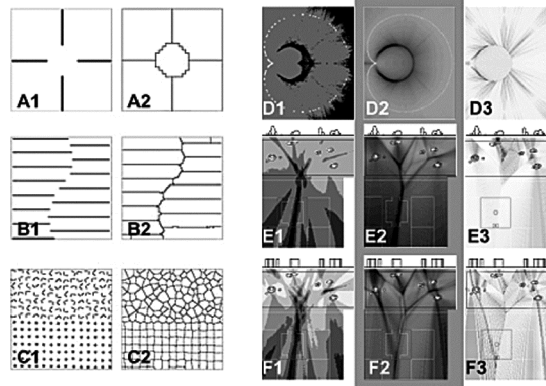


Figure 2. Ehrenstein illusion (A1), and its segmentation (A2). Similar data for illusory contour (B1~B2) and texton pattern (C1~C2). Adapted medial axis transformation (Van Tonder & Ejima 2003) of cardioid shape (D), Ryoanji (after Shigemori 1936) with rock height-related (E) and uniform (F) contour values. Computations were performed in noise resistant (D1, E1, F1), intermediate (D2, E2, F2) and noise sensitive (D3, E3, F3) modes.

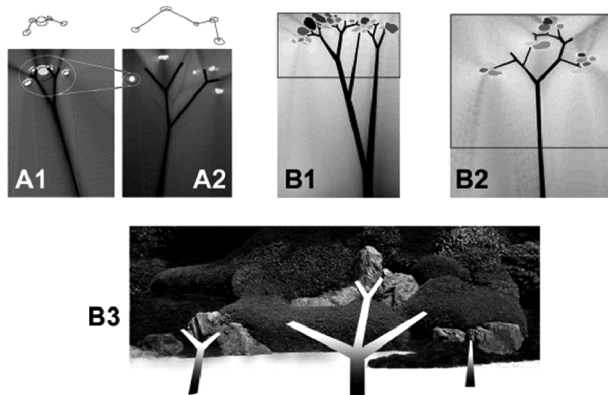


Figure 3. Local (A1) and global (A2) dichotomous tree structures in Ryoanji (Van Tonder & Lyons 2005). Examples of more intricate, strict dichotomous tree structures (B2) in karesansui also exist (Van Tonder 2006 in press). Non-dichotomous examples are common (B2~ B3), a consequence of design complexity.

medial axes do not significantly change, while the most prominent of the rock clusters in *Ryoanji* repeats the same structure, again converging onto the same viewing location (Figure 3A1~A2).

How does this subliminal structure possibly relate to our perception of the garden, and a sense of its order and complexity? Various theories and discoveries of brain mechanisms indicate that some parts of the visual system operate in

a predominantly object centered mode while other parts are mainly viewer centered (Biederman 1987; Milner & Goodale 1995; Tarr 1995; Tarr & Bülthoff 1995). From a mathematical point of reference, medial axes have great potential in representing shape (Blum 1973) and as a memory structure of the history of structural deformations that a shape has undergone (Leyton 1992). If the medial axis were truly used as internal abstract scaffold of shape, one could argue that the rotational transformations required to map between object and viewer centered representations of *Ryoanji* are minimized around the canonical viewpoint, since the viewer is aligned with a generic view of the global skeleton of the composition. The reduced perceptual effort may partially explain the sense of a transition towards visual calm when the visitor moves from the garden entrance towards a seat on the viewing verandah. This observation may also prove useful in understanding why humans naturally prefer to orient themselves relative to objects in a certain way.

Viewed from the entrance, the rock composition demands a large extent of optical accommodation as one's gaze shifts between nearby and distant clusters. The smallest shift in viewing position further causes a dramatic parallax shift between clusters, enhancing the sense of depth. To emphasize the effect, the largest rock in the garden is closest to the entrance, thus biasing the size gradient across the composition such that it is heightened at the point of entry. Around the canonical viewpoint, i.e. close to the central trunk of the medial axis structure, minimal adjustment of accommodation is required. From this vantage point the variance in distance between the viewer and each rock cluster is a minimum. When sitting on the verandah, as opposed to inside the temple, the garden is nevertheless still observed in a piecemeal fashion, as one turns the head from one direction to another as each cluster is viewed individually. The medial axis structure of individual clusters may become more dominant during such shifts, and are perhaps integrated into the global structure during such acts of visual scanning.

A hidden global structural hierarchy suggests that, just as one is able to immediately point to the invisible geometrical centre of a given shape, the viewer can in theory immediately evaluate the balance of the entire design and sense the structural order of the empty spaces. For example, the medial axis structure may or may not adhere to garden design guidelines such as the use of odd numbered visual junctions, triangularity, asymmetry, reduction of closure and avoiding straight paths. Figure 3B1 shows the mapping of a garden with an elaborate but surprisingly consistent dichotomous tree structure. As subliminal structure, the medial axis is most likely not intentionally created but results from deeply accomplished design intuition and true mastering of the classical guidelines. This is not the case with all gardens. First, because the intention with all designs is not to present a convergent empty space to the viewer. Dichotomous tree structures are especially prevalent in gardens of the *Myoushinji* sect of the *Rinzai* school of Zen Buddhism in Kyoto. Second, it may be due to the difficulty in mastering of this style. Figures

3B2~B3 shows examples where not all branches slope towards the viewer, where not all branching nodes are dichotomous, and where the overall balance of the tree is compromised.

The idea of a global structural hierarchy in *Ryoanji* enables a new level of comparison between *karesansui* and classical western gardens. Branching networks with centers of convergence, axial lines and predefined viewing locations are commonly found in both. We essentially found a subliminal visual center in *Ryoanji*. Arnheim's proposal that visual centers constitute the perceptual power (Arnheim 1988) in artistic compositions can thus also be extended to *karesansui* in a novel manner.

Is a structural arrangement such as *Ryoanji* accidental? The alignment between a subliminal structure and the architecture suggests that it is unlikely. As a test, we produced many randomized versions of the garden (Van Tonder et al. 2002, supplementary material), none which had similar characteristics (i.e. convergence onto the viewing point, consistent branching rule, roughly uniform distribution of branching nodes, balanced structural asymmetry, branches all sloping toward the viewer). In what follows here, I demonstrate calculation of the probability of accidentally recreating a structure similar to *Ryoanji*. Readers not interested in the technical details pertaining to this computation may skip the following section.

I have used combined independent probabilities to calculate a greatest likelihood measure of the probability of creating a similar dichotomous tree, in the empty space of the garden, by randomly tossing 5 rock clusters into a rectangle. First, the 20m $\times$ 10m rectangular envelope of the rocks of *Ryoanji* was represented by a grid of 20 $\times$ 10 cells, assuming that 1m² is a generous resolution for accuracy at this scale. The probability of any cluster c_i landing within any cell (x, y) is for all practical purposes mutually independent and equal for any given cell ($P(c_i/(i, x, y)) = P(c_i) = 1/200 = 0.005$ for $i \in [1..5]$).

Neglecting the small probability that more than one cluster may land in the same cell, the probability can then be computed as a string of unconditional probabilities, given that one understands the underlying conditional geometry of the convergent tree structure. Figure 4C1~C5 explains the geometry behind my reasoning. The two clusters closest to the viewing verandah determine the orientation of the main trunk of the medial axis tree, and thus the global alignment between the tree and the viewing point. Given that these two rocks (1) span the width of the garden, cluster c_1 should at least fall within one sixth of the left or right frontal section of the garden, and c_2 in the opposite half, i.e. $P(c_1) = P(c_2) = (7 \times 5)/200 = 0.175$. The average position of each cluster over many trials will thus be the centre of the respective shaded rectangles in Figure 4C1~C2. The entire composition should (2) face the viewer at a slight angle (see guideline 35 in Table 1), and hence the average location of c_2 should be behind that of c_1 . As an approximation I thus set it to the centre of the darker back half

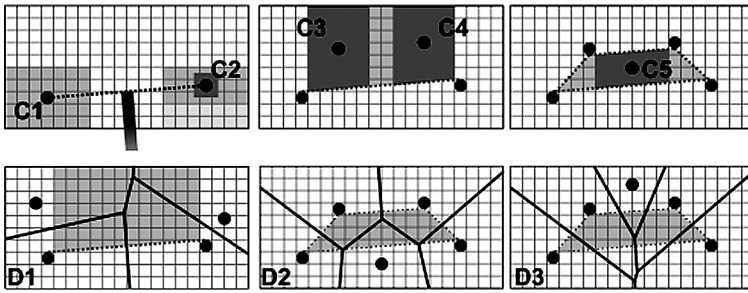


Figure 4. Probability of creating a dichotomous tree, converging within 1m radius of the main viewing point of Ryoanji is constrained by geometry of the front most pair of rock clusters, C_1 and C_2 , the back pair, C_3 and C_4 , and central rock cluster, C_5 . D_1 , D_2 and D_3 illustrate violations of these geometrical constraints.

of the shaded rectangle in Figure 4C2, and thus $P(c_2) = (7 \times 3)/200 = 0.105$. To ensure that (3) the medial axis between the two clusters is aligned to within 1m of the viewing centre, c_2 is restricted to a $2m \times 2m$ region around the centre of the small dark rectangle, so that $P(c_2) = (2 \times 2)/200 = 0.02$. For the upper branches of the tree (4) to slope towards the viewer, clusters c_3 and c_4 must fall behind and within the horizontal region delineated by c_1 and c_2 (Figure 4C3~C4), and thus $P(c_3) = P(c_4) = (12 \times 7)/200 = 0.42$. Violation of this condition prevents the formation of branches that slope towards the viewing area (Figure 4D1). For the sake of preserving the internal dichotomous branching rule and visual balance (5) of higher branches, clusters c_3 and c_4 should not intrude on the central band of the garden. c_3 and c_4 are therefore each restricted to one third of the width between c_1 and c_2 , and now $P(c_3) = (5 \times 7)/200 = 0.175$, $P(c_4) = (5 \times 6)/200 = 0.15$. Their respective average locations over many trials are indicated in Figure 4C3~C4.

The final cluster c_5 must fall within the trapezium (6) defined by vertices c_1 , c_2 , c_3 and c_4 , i.e. $P(c_5) = (10 \times 3)/200 = 0.15$. Figures 4D2~D3 illustrate how violation of this condition distorts the global convergence of the medial axis towards the viewing area and disrupts the internal branching hierarchy. For the sake of preserving the branching rule and balance of the tree (7) it must also fall within the strip between c_3 and c_4 , i.e. $P(c_5) = (6 \times 3)/200 = 0.09$. Conditions (5) and (7) are approximations derived from stricter limitations for preserving medial axis geometry. Note that the probability of creating a dichotomously branching empty space can also be computed by combining limitations on the internal geometry of each triangular set of rock clusters. This computation is not presented here, but interested readers may derive the result for themselves by using the geometrical lemmas which state that the three orthogonal midlines of each side of a triangle always (a) meet at a given point, p , in space, (b) where p lies inside the triangle only

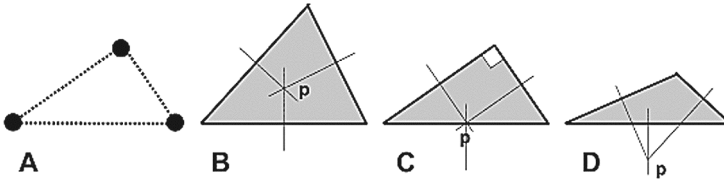


Figure 5. Understanding medial axis geometry in a triangular unit of objects (A) is simplified by observing that the midlines of each side of a triangle always intersect at a point p , where p falls inside an acute triangle (B), coincides with the longest side of a right triangle (C) and falls outside an obtuse triangle (D).

if all three vertices are spanned by acute angles, (c) coincide with the midpoint of the longest side of a right triangle and (d) falls outside the region delineated by an obtuse triangle (Figure 5). Note that, in *Ryoanji*, all except the backmost combination of three rock clusters delineate right or obtuse triangles where p lies between the observer and the triangle, such as shown in Figure 5C~D.

The average location of c_5 over many trials is shown in Figure 4C5. The probability of accidentally creating this structure is then calculated by combining each individual probability, as follows:

$$\begin{aligned}
 P(\text{dichotomous tree}) &\leq P(c_1 c_2 c_3 c_4 c_5) \\
 &= P(c_1)P(c_2)P(c_3)P(c_4)P(c_5) \\
 &= 0.175 \times 0.02 \times 0.175 \times 0.15 \times 0.09 \\
 &= 8.3 \times 10^{-6},
 \end{aligned}$$

a probability of creating eight dichotomous tree structures in every one million attempts to randomly cast stone clusters into the garden – perhaps not a large number for a computational simulation, but certainly restrictive during the actual construction of a garden. Combining the probability of creating a dichotomous tree within a rock cluster of five rocks (as is the case with the left most cluster in *Ryoanji*) considerably lowers the random gardener's luck. Generously speaking, a cluster would fit into at most one sixth of the garden space, i.e. 7×5 cells. The probability of randomly creating a dichotomous branching structure with five rocks, c_{1j} where $j \in [1..5]$, in a 20×10 grid amounts to each of the previous cluster probabilities scaled by $(7 \times 5)/200 = 0.175$, ignoring the stricter additional condition that the geometrical centre of the rock cluster is actually fixed. Thus the probability of randomly creating a dichotomous tree in the empty space of one rock cluster is:

$$\begin{aligned}
 P(\text{cluster tree}) &\leq P(c_1 c_2 c_3 c_4 c_5) \times 0.175^5 \\
 &= 1.4 \times 10^{-9}
 \end{aligned}$$

or roughly once in every thousand million attempts, and the chance of accidentally creating a version of *Ryoanji* with both a local and global dichotomously branching empty space structure is:

$$\begin{aligned} P(\text{Ryoanji}) &\leq P(\text{dichotomous tree}) \times P(\text{cluster tree}) \\ &= 1.1 \times 10^{-14} \end{aligned}$$

The random gardener at best stands a chance of creating the structural semblance of *Ryoanji* only once in every one hundred million million attempts of enthusiastically casting stones into the void. This value is only a best likelihood measure. With more accurate geometrical constraints the probability would be even smaller. Without the well worked out design guidelines of the Japanese gardening tradition such a non-accidental structure would not have been created.

Depiction of visual depth in dry rock gardens: From illustrations to reconstructed garden maps

Depictions in parallel perspective appear so often in Japanese (influenced by Chinese art) painting that it is considered the representative system of depicting depth in Japanese illustrations. One possible motif for the use of such a perspective system is that it allows one to browse through a scroll, shifting from one scene to the next without presenting awkward transitions in depth perspective.

A search for maps of gardens no longer in existence cast the topic of depth depiction in a new light (Van Tonder 2006 in press). If the parallel perspective system were followed consistently in a garden illustration, one should be able to reconstruct the map of that garden in a relatively simple spatial transformation. Figure 6 shows three basic perspective systems. A map (A) of the grid depicted in parallel perspective can be obtained by transforming an appropriate parallelogram in B back into a rectangle. Linear perspective (C) calls for mapping of a trapezium onto a rectangular map, while vertical perspective (D), where objects are stacked one on top of the other in fronto-parallel planes, require only a change in aspect

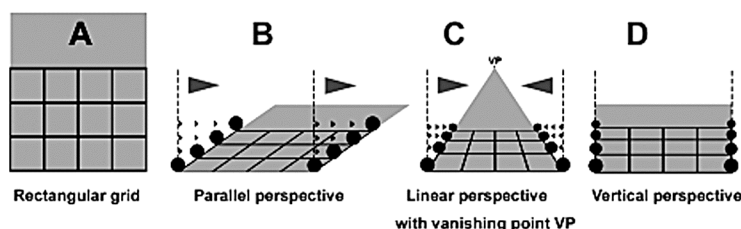


Figure 6. A rectangular grid (A) depicted in parallel (B), linear (C) and vertical perspective (D).

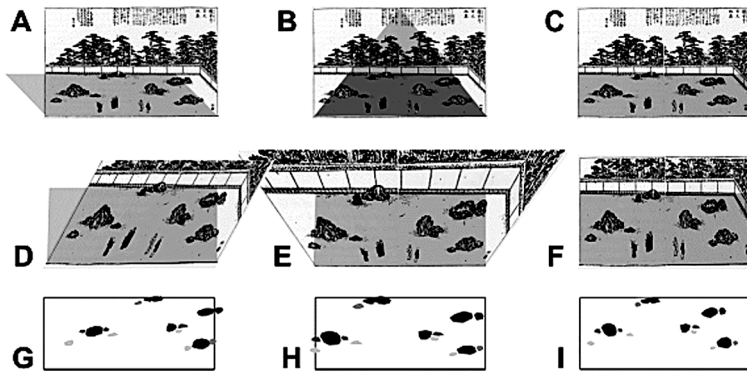


Figure 7. Shaded regions for reconstruction of the map of Ryoanji from an illustration (A) by Akisato (1799) assuming parallel (A), linear (B) and vertical (C) perspective, and after the required spatial transformation into a rectangle (D~F). Locations of rocks clarify the resultant maps (G~I).

ratio of the depicted scenery. Each perspective system can be described in much greater detail (Panofsky 1927; Kemp 1990).

Figure 7 shows the reconstruction of maps of *Ryoanji* from an illustration by the Edo-period (1603–1867) garden connoisseur Akisato Ritoh (Akisato 1799, 1828). The top row in Figure 7 shows the shaded regions that should be transformed into a rectangle if the garden were depicted in parallel (A), linear (B) and vertical (C) perspective. Assumptions for the choice of the relevant shaded regions appear elsewhere (Van Tonder 2006 in press).

The middle row (Figure 7D~F) shows the shaded regions after transformation, and the bottom row shows the resultant maps of *Ryoanji* that one would obtain (Figure 7G~I) given the assumption of each system of depth depiction. Given the predominant use of parallel perspective in Japanese painting, one would expect Figure 7G to most closely resemble the actual map of *Ryoanji*. To the contrary, and this turned out to be the case with various other illustrations as well, reconstruction in vertical and linear perspective far outperform that of parallel perspective, with linear perspective usually a little less accurate than vertical perspective. Accuracy was measured as the variance between rocks in actual and reconstructed maps, as explained in Van Tonder (2006 in press). But the result holds true only for the rock compositions of the gardens. Actually, reconstruction in parallel perspective was clearly superior with respect to depicted rectilinear structures, such as courtyard walls, roofs, and verandahs.

Reliable reconstruction of maps is useful when dealing with gardens of which the original composition has not been documented but significantly altered somewhere in its history. For example, Akisato's illustration of the garden at *Zakkein* in Kyoto clearly shows that the design was dramatically changed at some point dur-

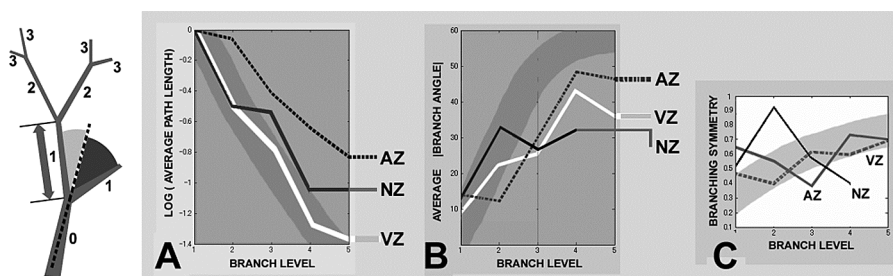


Figure 8. Comparison of average path length (A), branch angle (B), and branch symmetry (C) for three versions of *Zakkein* garden, Kyoto. In each case, dark bands indicate corresponding tendencies in natural dichotomous tree structures. Branch level 0 denotes the main trunk of a tree (far left).

ing the last 200 years. Comparing branching properties (Figure 8) of the medial axis in the vertical perspective map VZ reconstructed from Akisato's illustration, a reconstruction NZ by Nakane (1999), and the altered version AZ of *Zakkein*, the reconstruction VZ most closely approximates structural properties, namely average path length (Figure 8A), branch angle (Figure 8B) and branch symmetry (Figure 8C) observed in naturalistic branching structures (dark bands in Figure 8A~C). The graphs caution against the alteration of an existing masterpiece. In the case of *Zakkein* the medial axis tree obtained from VZ is of exceptional structural complexity – a structure that would probably require more than just one hundred million million random casting of stones into an open courtyard. Sadly, this garden was completely demolished in recent years. Created more than 400 years ago, *Zakkein* belongs to the same greater temple complex as *Ryoanji*. This garden could thus provide important clues to resolve disputes (Kuitert 2002) regarding origin, style and identity of the anonymous creators of gardening masterpieces such as *Ryoanji*.

Results from the described reconstruction technique have interesting consequences for an understanding of the depictive space in Japanese illustrations. First, the reconstructed maps suggest that naturalistic objects comfortably map into two dimensional depictions, while the depiction of rectilinear shapes are riddled with difficulties. Second, a composition of natural objects as a whole can be more easily be depicted in two dimensions. In other words, the spatial relations between natural objects are also more accurately depicted.

Regarding the first statement, given the apparently amorphous shape of rock, this difference is to be expected. Human subjects prefer generic views of objects (Palmer, Rosch, & Chase 1981). Thinking of the *viewing sphere* (Koenderink & Van Doorn 1979) of effectively amorphous, convex-hull shapes of the triangular trees and rocks preferred in *karesansui*, it is clear that naturalistic objects would

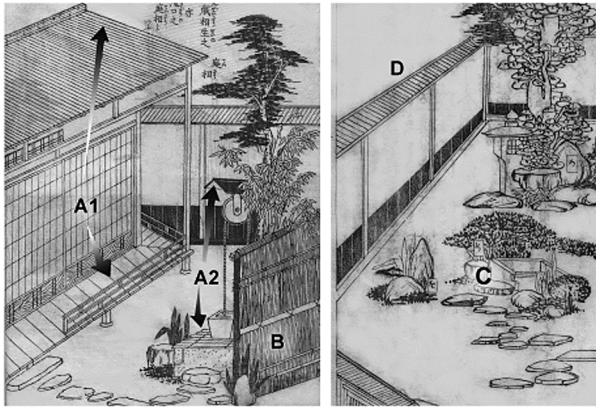


Figure 9. Far from consistently using parallel perspective, Akisato Ritoh (1828) often depicted different parts of the same shapes in conflicting (A1~A2), inverse (B), isolated (C), and rarely, linear perspective (D).

not significantly deviate from their generic view as the viewing location change. But small changes in viewpoint have large influences on the visual appearance of a flat surface. The visual appearance of a wall, which remains stable only over a narrow range of angles, will deviate dramatically from its generic view even for small changes in viewing location of the human subject. Depiction of such objects does not naturally appeal to the artist, as can be seen from various illustrations by Akisato (1828). In the same picture, the edges of the bottom and the top of structures such as wells and houses are presented along different parallel perspective lines (Figure 9A1~A2), in inverse perspective (Figure 9B), isolated perspective (Figure 9C) or, in rare occasions, even in uncannily correct linear perspective (Figure 9D). I would like to suggest that, at least to some extent, opposing perspective was an artistic device in itself, used to create a more interesting depicted space (Kishi 1994), or to improve visual balance of the composition. Opposing perspective could also have resulted from the complexity of drawing in an overall consistent perspective system. The natural tendency in which humans draw pictures have been compared to a spatial laying out of consecutive canonical views of the object (Willats 1997). This tendency seems to dictate the use of two different perspective systems in Akisato's art when he deals with man-made and naturalistic objects – an observation that may hold true for Japanese traditional painting in general.

The second issue relates to the significance of a global structural scaffold, such as the medial axis tree shown previously. The fact that it converges towards the viewpoint, and the care taken during garden construction to ensure that the most desired visual features of each rock faces toward the same viewing location suggest that neither individual rocks or the composition as a whole can actually be equated

to an amorphous convex hull. Rather, the canonical viewpoint in such carefully tuned compositions signify a special instance where the shapes of each part, their relative spatial arrangement and the global structure are all seen in their generic views. All objects are already presented in a pseudo flat pictorial space. Unlike fronto-parallel planes, objects are here presented in canono-parallel surfaces, concentric around the canonical viewing point. Hence the two-dimensional depiction of such a multi-scale generic presentation should therefore come more natural than a combination of walls and verandahs, where one can hardly even imagine how each individual object and the structure as a whole could be presented in one all-inclusive generic view. Interestingly, Akisato depicts most gardens from their canonical viewpoint (known from factual information regarding the architectural style in which many of the gardens were originally constructed). Visualizing *kare-sansui* from unusual viewpoints first appears after the invention of the camera, when photographers looked at gardens through the frame of the view finder, or rather, were faced with the dilemma of fitting the garden into that frame.

In Arnheim's definition of order, this kind of design therefore features hierarchical, coordinated and accidental orders. In his definition of complexity, these gardens are therefore not only highly ordered but also highly complex – the hallmark of a masterpiece. The combination of rectilinear and natural objects in Japanese landscape design is noted for its ability to evoke a strong sense of space and visual interest, even in comparatively empty settings (Nistchke 1966, 1993). The addition of a rectilinear system of order to an already multi-ordered naturalistic rock composition should thus further increase the complexity of the design. I would like to speculate that such a highly ordered and complex visual composition, with both naturalistic and rectilinear components, provides a rich stimulus to both the parietal and temporal pathways in the brain.

Visual completion in dry rock gardens

The observer's experience of visual complexity in *karesansui* derives to a large degree from active perceptual reconstruction of incomplete visual cues. Occlusion junctions mentioned in the previous section constitute just one example where the visual system must extrapolate from local parts to reconstruct the more holistic structure of the garden composition. *Karesansui* engages the viewer in perceptual completion of the incomplete via a multitude of design effects.

Many of the strict specifications regarding placement of rocks can be interpreted as guidelines to ensure visual junctions at right or acute angles between rocks and gravel (Figure 10A). Junctions at obtuse angles would occur when a rock is buried shallowly (Figure 10C). Instead of grouping with other garden elements, perceptual completion would reconstruct the bounding contour of the rock as a separate, enclosed shape (Figure 10D). With appropriate junctions a rock

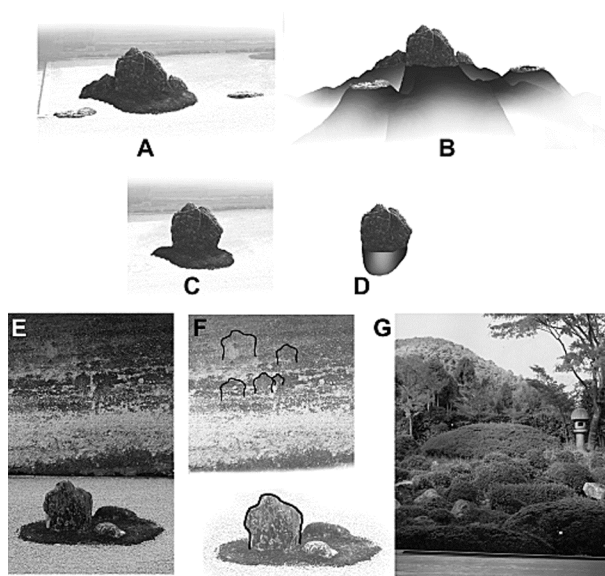


Figure 10. Depending on visual junctions between rocks and gravel (A) a cluster would complete as a part of a more extensive mass (B) or pop out (C, D) as an isolated object. An earthen wall may provide the ideal rough texture within which the rock composition can be perceptually completed (E). For example, groupings similar in shape to that of rocks may become salient in this texture (F). In *shakkei*, the partial view of a mountain (G) engenders complex perceptual completion and grouping between garden elements and the outside landscape.

appears to diverge below ground. Divergent junction angles steer the process of completion towards incorporating the figure into the larger structural frame of the invisible underground bedrock. Moss and smaller rocks placed around the foot of main rocks further help to correct and diversify the range of possible completion junctions for the incomplete shape of the rock cluster.

The use of homogenous natural textures further aids mental completion. If the rock composition were to convey a sense of naturalistic order and complexity, it should consist of a much greater number of components strung together in a continuous structure, to more closely approximate the fractal structure of real geological formations, such as a mountain range. Visual perception is obviously free to complete the whole from very few islands of structure. But in *karesansui* perceptual grouping mechanisms are provided with a rich palette of potential shapes, in the form of rough textures in gravel and earthen walls (Figure 10E), from which the limited store of rock clusters can be extended as if it were an infinitely extending mountain range (Figure 10F). I can hardly imagine a more effective means



Figure 11. Cherry blossoms at *Kiyomizu* temple. Ai Akino. *Nihonga* pigment on cedar panel. Emma votive tablet at the inner precinct of *Kiyomizu* temple, Kyoto 2004.

for suggesting greater lawfulness and multifariousness (after Arnheim's definition of order and complexity) in the internally perceived relations between parts and wholes.

The incomplete is further encountered in the form of objects partly occluded by courtyard walls. In doing so, for example by showing only a few branches of a tree, the viewer is invited to reconstruct the tree as a perfect whole according to her internal model of the tree. *Shakkei* (borrowed scenery) is a well known design technique to draw external components into one's perception of the garden space. A common example of *shakkei* involves the inclusion of a partly occluded view of a distant mountain (Figure 10G). With only the mountain top visible in the garden, the perception of the whole is affected to the extent that the garden appears more spacious than it actually is. Rock clusters may seem to extend towards the mountain, the latter appearing much closer than it actually is, and courtyard walls often appear much lower than they actually are. A thorough discussion on *shakkei* justifies an entire book dedicated to the topic.

Japanese painting brims with examples where shapes are only partly visible. Thus the observer can sense the whole according to her own internal perceptual makeup, and sense a world extending beyond the frame. One appropriately presented branch of a flowering cherry tree has the power to evoke a clear sense of a huge flowering tree beyond the confine of the viewing space (Figure 11). By presenting such islands of structure, the artist deals with order and complexity at a sophisticated level. Individual parts should be more manageable in terms of complexity. The choice in how much order to install in their combination is thus the artist's main challenge. A design left sufficiently open-ended allows viewers to reconstruct visual relationships according to their own capabilities and internal wiring. Present research on human perception of fractal dimension and other structural properties may have to consider internally completed perceptual structure before the true role of fractals etc. will be more deeply understood.

Trees

According to the illustrated garden manual on mountain and water landscaping, *sansui narabini yagyō no zu* (Shingen 1466) rocks and trees constitute the structural backbone of a garden. The latter half of this chapter looks at design with trees, specifically how Japanese gardeners and *ikebana* (Japanese flower arranging) artists shape arboreal material, and the perceptual effects achieved. It is not as if rocks and trees can be separated without sacrificing some sense of holism. The wealth of issues pertaining to plants and trees in particular is fortunately sufficiently numerous to justify a section devoted to the role of trees in shaping a garden.

Similar to painted trees, garden trees and *ikebana* materials share suggestive depictive qualities. Accomplished designers intentionally present the viewer with a canonical view of bilaterally asymmetric but visually balanced tree shapes. Branches are uniformly reduced in structure, to decrease occlusion of background objects and selectively reveal segments of the main trunk, branches, bark, clumps of leaves and individual leaf shapes. One finds many similarities and some fundamental differences between the arrangement of rocks and trees. Both, however, share a strong depictive aspect.

Simple computer generated trees and some stylized depictions of trees in scrolls are used in a first approach to gain deeper understanding of the impact of Japanese garden aesthetics on tree shapes. Formalization of tree structure for this purpose benefits from two well known models regarding structure and context generation in literary bodies (Zipf 1949; Simon 1955) and music (Zanette 2006).

It is hard to imagine that any seeing human being would never have set eyes upon the intricate network of branches of a tree, whether in the wilderness or in a garden. The naturally balanced shape of an undisturbed tree contrasts starkly with the ripples of confusion evident where human hands have impinged. Thus, ideally, real natural trees and real human shaped trees should form part of this investigation. Difficulties with data collection from real trees prohibited an in depth analysis at this point. Computer generated trees intrinsically obey the same laws as their linguistic and musical counterparts, casting the visual backbone of the garden in a new light. Perhaps music and literary works need reinterpretation within a visual framework.

The next section will recount some conceptual guidelines by which Japanese gardeners create a sense of visual harmony in the trees of their tamed landscapes, from which a core of characteristic structural features of garden trees is extracted. Models of perceptual structure in language and music are then applied to tree structures.

Classical instructions regarding the arboreal backbone of the garden

The most fundamental of *Zouen's* instructions apply to both rocks and trees: each object should be conceived as a triangular building block with a dominant line of orientation. Thus trees should be chosen with the emphasis on vertical, horizontal and diagonal lines (Figure 12), representative of the relationship between the divine, earth and man. A grouping of rocks and trees can be considered inarticulate without expression of this triad.

At the time when *Zouen's* teachings were transmitted (Shingen 1466), the practice of collecting garden trees from the wild was still common. Just as rocks were chosen based on textural grain and shape silhouette, so were trees selected for 'interesting and graceful' shape. Various aspects of tree shaping are presented and discussed in the classical manual. Classical design guidelines regarding plants are listed in Table 2 along with my interpretation of the relevant Gestalt factors, depictive effects, and relevance in terms of the order and complexity of the design.

The importance of the diagonal, vertical and horizontal forces is emphasized in various passages in the text. *Zouen* describes how one can effectively depict different types of scenery (e.g. seashore, craggy mountains, river valleys, etc.) by carefully matching each type of tree with its exact geological setting and natural habitat, paying attention to the trees and other plants it is normally associated with. Other scenic effects are also discussed, such as 'blocking the path' by grouping a few trees, contorted such that their trunks cross one another, 'like parents watching over their child' (Figure 12D). Such effects find their counterpart in the guidelines concerning rock clusters. For example, folding screen rocks, evoking crags jutting out one in front of the other, is the equivalent of 'path blocking'. Such



Figure 12. Trees with (A) vertical, (B) diagonal and (C) horizontal trunks. Upper tips of branches should slope vertically. (D) Path blocking trees. Shingen's manual, transmitting the teachings of 11th century gardener priest *Zouen*, depicts mainly pines grouped with rocks. Note the triangular shape of leaf clusters on every branch, and of each tree as a whole.

Table 2. Classical Japanese design instructions on shaping and applying arboreal materials in the garden (column 1) is related to Gestalt grouping factors (column 2), depictive effects (column 3) and Arnheim's conceptualization of order and complexity (column 4).

Design guideline	Gestalt factors	Depictive effect	Order and complexity
1 Regard (rocks and) trees as the backbone of a garden composition.	All	Global structural scaffold during construction of garden.	Clarify the orders. Reduce complexity.
2 Trees with distinctly horizontal, diagonal or vertical sloping trunks are preferred.	Similarity, asymmetry.	Increase pictorial quality of trees, while regularizing visual structure.	Increase order. Reduce complexity.
3 Even if the trunk of a tree is horizontal or diagonal closer to the ground, the tips of higher branches should eventually slope along a vertical direction.	Similarity at highest spatial scale.	Unification of visual structure in given domains of the design. Simplification of visual texture.	Increase order at higher spatial scales. Increase homogeneity. Clarify structural hierarchy. Reduce complexity in the sense that the multifariousness of orders is reduced.
4 Trees must be positioned to cause minimal visual obstruction of other objects.	All factors. With more layers of visual structure made visible, perceptual grouping is more complex.	Enhance the contrast between convergent and divergent visual structure.	Appropriate arrangement of objects increase order in design. Complexity is enhanced by increased range of visible orders.
5 The tree should be pruned to have as few branches as possible, but with great sensitivity to the balance of shape.	Simplicity of shape, clarifying self-similarity of tree shape, while increasing visibility of more layers of visual structure, thus affecting perceptual grouping in all aspects. Bilateral asymmetry.	Structural layers enhance the sense of looking into a fronto-parallel pictorial space. Objects usually face the viewer from a predetermined direction. This guideline ensures access to such generic views.	Clarifies hierarchy, thus increases order. Reduce the number of instances of order, thus decrease complexity. Increase salience of substructure in remaining branches, thereby enhancing different kinds of order and thus increasing the complexity.
6 The ends of branches must not be too sparse nor left bare.	Increase self-similarity, induce good continuation. Increase closure at very large and very small spatial scales of visual structure in the tree.	Increase number of occlusion layers in the visual scene. Enhance percept of spatial continuation, spatial depth and looking into a pictorial composition of fronto-parallel structural layers.	Enlarge hierarchy and range of orders. Increase complexity.
7 Trees with fewer lower branches but graceful forms are preferable.	All factors. Similar to 4 and 5. Proximity (prevent crowding of visual structure)	Provides complete occlusion of one structural layer by another. A transition from one visual texture into another, it prevents salient line ends from drawing one's attention to associations of dead biological structure.	See #5.
8 Gnarled branches should not be removed.	Preserve structures rich in multi-scale similarity. Increase bilateral asymmetry. Reduce closure and good continuation.	Enhance the percept of looking at fronto-parallel layers in pictorial space. Provides a degree of self-similarity, usually obtained only over many more spatial scales, within non-occluding structure, thereby enhancing the percept of looking at a fronto-parallel pictorial layer.	Preserve dense source of hierarchical structure and accident. Increase visual complexity.

Table 2. (continued)

Design guideline	Gestalt factors	Depictive effect	Order and complexity
9 Lower branches should be left in abundance for species prized for their foliage.	Increase closure. Simplicity, self-similarity (considering the kinds of trees prized for foliage, such as the Japanese maple, <i>Acer japonica</i>).	Enhances textural (pictorial) quality. In some cases, leaves may be so delicately coloured/shaped as to appear real, but rather painted.	Increase closure. Simplicity, self-similarity (considering the kinds of trees prized for foliage, such as the Japanese maple, <i>Acer japonica</i>)
10 A tree should give the impression of 'ten thousand trees in a single glance.'	Similarity over multiple scales.	If the entire tree shape repeats in branches, and if branches visually alternate, leaf clumps of the tree resemble woodland scenery.	The co-presence of orders occurring at different spatial extremes enhances both hierarchy and coordination. Increase complexity.
11 Where more than one branch point in the same direction, or when branches are too close together, the weaker branches should be removed.	Reduce similarity and symmetry. Adjust proximity.	Avoid unnatural shape, improve uniformity of visual density of structure, and remove excessive local bilateral symmetry and similarity.	Reduce disorder (clash between parallel with non-parallel branches). Remove simplistic order. Increase complexity.
12 One should avoid branches pointing directly towards the viewing location.	Simplicity. Avoid closure.	Avoid non-generic object views and arrangements that are convex toward the viewer.	Remove simplistic order and complexity at higher spatial scales, increase order and complexity at larger spatial scales.
13 While dead or drooping branches are not a taboo, they should never be in the vicinity of, or point towards, the viewing location. Drooping branches are acceptable only for weeping tree shapes.	Avoid closure.	Avoid shapes that are convex (and thus exclusive) towards the viewer. Avoid visual metaphors of bad health or gravitational load.	Similar to #12.
14 Exotically coloured species should be well hidden among other trees	Increase simplicity and similarity.	Increase homogeneity, monochromaticity and pictorial quality of the design.	Reduce order and complexity.
15 Trees should be chosen according to the natural habitat in which they occur: Mountain trees for mountainous slopes, shade loving trees for deep valleys, domesticated trees for scenery suggestive of villages.	All. A complex aspect of grouping, involving more abstract learnt categories and their associations with non-visual attributes.	Sound imagery of natural ecology makes the garden more convincing as a depiction of a naturalistic space that should be much larger than the actual courtyard.	Increase coordination. Reduce complexity.



Figure 13. Shrubs in the *Shisendo* garden, Kyoto, are arranged into a folding screen formation. This technique enhances spatial depth perception.

methods seem to enhance the sense of spatial depth. Shrubs, which are more solid in appearance, are arranged similarly into formations of visual overlap (Figure 13).

Practices of the present generation of Japanese gardeners

Collaborating with garden designers presently active in Kyoto proves that many of the old guidelines still provide a foundation in creation, alteration and maintenance of gardens. In addition, I found the following concepts presently used by gardeners relevant to a better understanding of tree shaping:

- Evenly spread branches improve physical health and visual balance of the tree.
- When pruning, one also visualizes changes to the root system underground.
- Hard unfertilized soil slows tree growth, increasing the hardiness and lifespan of the tree. Slow growth improves visual balance and reduces intensive pruning.
- Trees are always growing, and one has to anticipate how pruning in the present will affect the shape much later. For example, even a beautifully shaped branch may have to be sacrificed at some point as the tree becomes larger, to make space for a more suitably proportioned branch at the appropriate time.
- One should be particularly aware of how one branch overlaps with another.
- The tree as a whole and each of its branches should have a *soft* appearance.
- Correct pruning selectively exposes the ‘flesh, skin and bones’ of the tree.

Each statement is related to perceptual grouping. Evenly spreading of branches regulates density and proximity of parts; visualizing the underground root structure extends conscious awareness of asymmetry, closure and good continuation; slow growth allows fine tuning of structural properties, specially the density and proximity of parts and the multi-scale similarity of parts; sacrificing a branch at some point may restore the proximity and similarity of parts; overlap of branches is a decisive factor in tuning the degree of good continuation, asymmetry, proximity and similarity between parts. The resulting depictive effects follow suit from the examples listed in Tables 1 and 2.

Even if INTUITIVELY grasped, ‘softness’, and ‘flesh, skin and bones’ defy concise definition. I translate them according to discussions and participation in

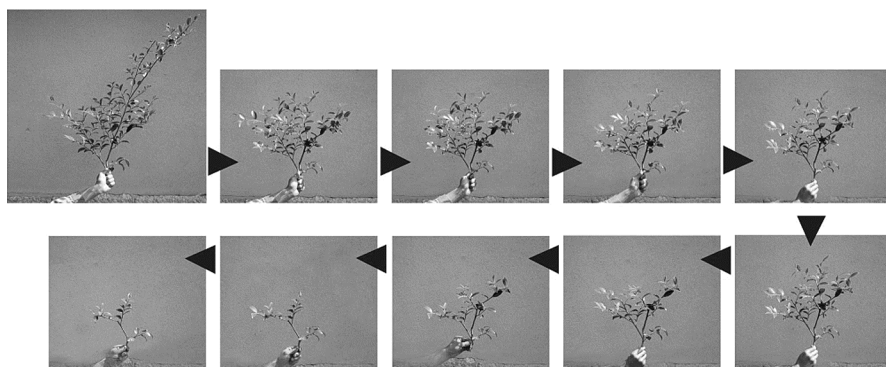


Figure 14. A Sazanqua branch trimming sequence shows the order in which parts are removed. The gardener gives priority to overall symmetry, visual balance, protrusions, uniformity of the density of twigs, and alternation of twigs along the primary branch.

actual pruning sessions with two master gardeners in Kyoto. *Softness* of appearance relates to the outer boundary contour of the tree silhouette and of individual branches. When pruned, the region furthest away from the trunk should always be defined by clumps of leaves, without any pruned stumps visible. Where stumps inevitably point toward the viewing direction, the angle of the cut surface is carefully chosen to face away from the viewer, thereby limiting the salience of the otherwise striking white cut mark (due to underlying fresh wood). If necessary, the cut can be painted the same colour as the leaves, since in their words ‘the human eye is most easily convinced by colour’. *Softness* further refers to the degree of occlusion, determined by the density of leaves. A carefully trimmed branch will have less solid clumps of leaves. As a result, the entire branch would appear lighter, more feathery in structure. It would react more visibly to delicate changes in the environment, such as a light breeze, and hence literally be suggestive, or depictive, of qualities usually perceived in soft, feathery structures. A *soft* branch thus also follows the classical guideline of lessening the number of parts on a tree and of reducing the occlusive impact of the tree. *Softness* in this regard also refers to the textural equivalent of transparency, a quality dubbed *textural laciness* in visual psychophysics (Watanabe & Cavanagh 1996).

For practical demonstration I recorded pruning sequences where a skilled gardener systematically pruned away the limbs on a branch while maintaining its softness and balance (Figure 14 shows an example). The density of smaller branches remains uniform throughout. The silhouette of the branch first increases in bilateral symmetry around the main trunk, and eventually becomes more asymmetrical again as very few branches remain. The degree of asymmetry of the network of smaller branches remains constant, however. At the last stage of the session, the gardener insisted that he could not go any further without completely destroying

the shape of the branch. Note that at this point it is a simple dichotomous tree with three clumps of leaves – the classical structural unit of the old gardening manuals. Future work in this direction will focus on a tree pruning graphics interface to create a more extensive pruning sequence database, from which the process of naturalistic pruning may be formalized.

Flesh, skin and bones refer to a level of pruning where parts of the trunk and limbs are selectively covered by bunches of leaves. As a result, patches of bark or individual leaf shapes would attract attention in some segments of the tree, whereas the shape of leaf clusters or twists of the trunk and branches would visually dominate in others. This makes for dramatic viewing of tree shapes (Figures 15 and 16). I often notice that leaf clusters never appear stacked vertically. When placed horizontally or along a slight diagonal, the leaf clusters seem to slide around as one moves, the effect of parallax motion. This dramatic effect resembles sliding doors revealing parts of branches and the trunk, thereby also enhancing the sense of spaciousness and spatial depth in the garden, and intensifying the sense of looking *into* the garden.

I believe the technique emphasizes different spatial scales within the same visual structure. When *Zouen* recounts the old saying of ‘ten thousand trees in a single glance’ it possibly refers to the effect of showing flesh, skin and bones, as well.

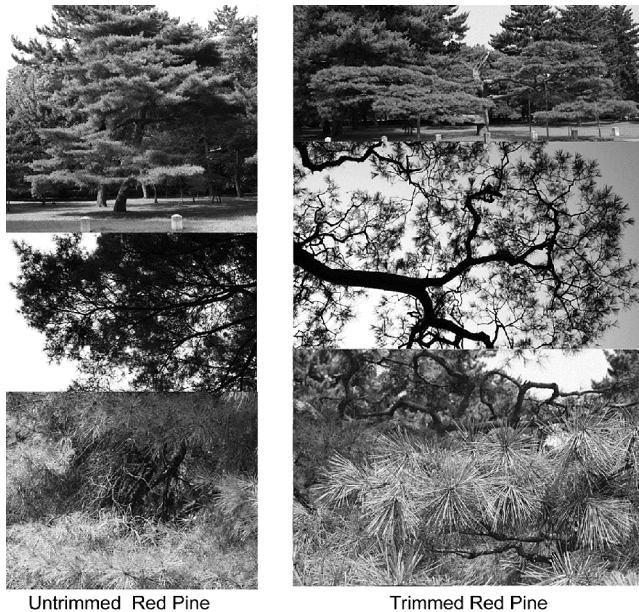


Figure 15. Different examples of Japanese red pines, one untrimmed (left) and another, carefully manicured over many years. Note how the trunk, branches, twigs and clumps of leaves are accentuated in different parts of the tree.



Figure 16. A close look reveals that various branch shapes repeat at small and large scales. Individual branches not only resemble entire trees; their combination gives the impression of a grove of trees on an incline.

From personal experience it seems that different branches within the same well shaped tree conjure an image that goes beyond a single tree. The branches themselves resemble different trees growing in a wooded mountain slope (Figure 16).

The intricate shapes of trees and rocks complicate garden design. Rocks are selected based on the harmony between surface grain, silhouette and volumetric shape. One is bound to present surface markings and rock shape in a natural attitude if the rock is placed as it was found in nature. Appropriate grouping with other stones further enhances this sense of naturalness, just as small parts of a fractal bear a specific resemblance to the global structure of the same fractal. Similarly, appropriately shaped clumps of leaves and their associated branches can evoke the shape of an entire tree, and even an entire grove of trees.

Gardeners do not explicitly articulate the principle, but repetition of shape over different spatial scales is a corner stone of *karesansui* design. Implicitly, the emphasis on triadic compositional units and diagonal, vertical and horizontal orientations voice the awareness of design at multiple scales. Triangular units lend themselves for incorporation into consecutively larger triangular clusters to the point where relatively few objects have the potential to evoke many possible structural groupings while maintaining compositional balance. Seen in isolation, each triangular unit reveals its visual balance. Even if structural units do not follow strict triangles, the conceptual framework simplifies the task of incorporating units into the larger composition without sacrificing visual balance. At a maximal scale in nature, mountains and trees usually approach triangular (actually trapezoidal) convex hull shapes. Relatively small scale triadic compositional subunits therefore also have the power to conjure multiple spatial scales of landscape scenery.

Nature abounds in examples where environmental challenges, such as poor soil, harsh climate and competition for resources drive organisms into tightly knit ecological communities with stunted growth, where plants reach only a fraction

of their true mature dimensions while slowly growing over a large number of seasonal cycles. The result is the arboreal counterpart of the slowly weathered rocks preferred in Japanese garden design. Resembling fully grown trees of great age, the miniature visually appears gnarled and twisted at every spatial scale, evoking a range of spatial impressions while physically remaining small enough to comfortably fit into a confined space. And with *miniature* I refer not only to prototypical small *bonsai*. A 5 meter version of a 50 meter tree is also a valid miniature. The fundamental visual difference between mature and miniature forms is that in the latter the global shape silhouette, the trunk and individual limbs, and the shapes of leaves and surface markings of the bark are visible in the same glance without the need for large shifts in gaze and visual accommodation, whereas one either sees the global shape of the mature tree from a distance with some loss of finer spatial detail, or the finer details are emphasized at closer proximity, for the sake of which the view of the global shape is sacrificed. Inspection of such a large object naturally also requires significant visual accommodation for each part observed. This may partly explain the coveting of the miniature in many cultures.

Shaping of branches in ikebana

Shaping of arboreal materials in gardens and in *ikebana* (Japanese flower arrangement) greatly overlap. Both are characterized by an emphasis on triangularity. The main difference lies in the fact that future growth of branches used in *ikebana* does not have to be considered. Flower artists can apply various methods in addition to pruning to obtain desired shapes. In my experience the most dramatic effects are achieved through the patient yet energetic plying and bending of material. A branch may be bent hundreds of times until its lines conform to the three ideal orientations. Instead of only removing branches or trimming out leaves and twigs, the artist literally reorganizes the spatial distribution of branches. Limbs on a trunk can be bent and twisted until they are more evenly distributed in three dimensional space (Figure 17). Leaves may also be repeatedly stroked and stretched to obtain a more spread out appearance in agreement with the overall triangular shape.

Viewing direction is of great importance in *ikebana*. Care is taken not to let any branches point toward the viewer, for reasons listed in Table 2. The viewer is presented with a canonical view of the composition as a whole and of each subpart, similar to the composition of gardens such as *Ryoanji*.

Formalizing the structure of artificially shaped trees

The following list of structural properties characterizes manicured garden trees:

- Bilateral asymmetry of the global tree shape
- Regularization of the shape envelope of individual branches

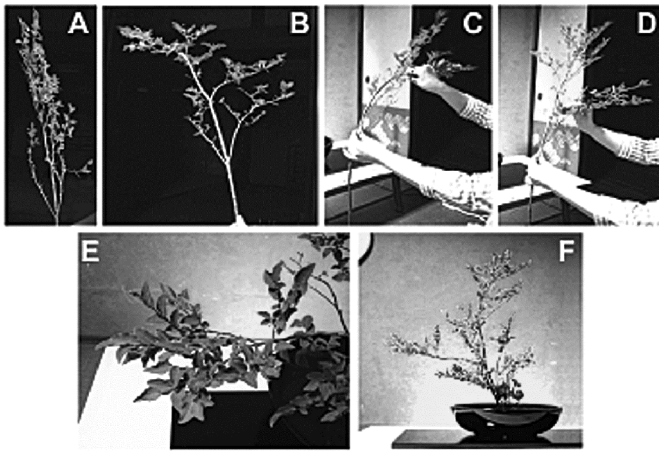


Figure 17. A sheet-like branch of natsuhaze (A) is twisted and bent until it assumes volumetric triangular facets (B). The branch as a whole (C) and individual side twigs (D) are bent to regularize their spatial distribution. (E) Even leaves are bent. The resultant arrangement (F) is a complex balance of triangular groupings.

- Balance of shapes at all spatial scales
- Repetition of shape over different spatial scales
- Compression of the number of spatial scales over which structures repeat
- Presentation of a canonical view of the tree
- Presentation of a canonical view of branches
- Sparseness of structure, especially of primary branches
- Absence of abrupt terminations in branches
- Uniform distribution of parts
- Division of tree into patches with distinct visual textures, related to structure at different spatial scale

A deeper understanding of universal qualities of good tree shape does not follow suit from the above list of properties. Ideally, the list would result in a formal analytical approach for this purpose. Computation of fractal dimension of trees may prove useful in characterizing tree shapes (Mandelbrot 1977), but its shortcomings will perhaps be evident from the following section.

The author is lead by parallels between poetry, literature, music, and visual rhythms of the garden, as discussed in *sakuteiki*, in the search for an appropriate framework of numerical analysis. Two models considered have already been applied with success in analysis of structure of language and music. They are Zipf's law of word incidence (Zipf 1949) and Simon's model of the rate of context generation in literary bodies (Simon 1955).

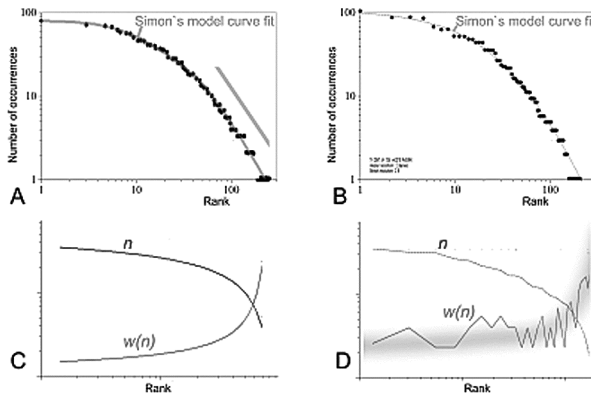


Figure 18. (A) Simon's model (red curve) accurately fits Zipf plot data points of word incidence against rank in an hypothetical literary work. Branch count (B) of a simple computer generated tree structure displays a similar trend. (C, D) Zipf variables n and $w(n)$ display a characteristic inverse relationship.

In a nutshell, Zipf's law states that the number of words $w(n)$ which occur exactly n times in a literary corpus varies with n (Figure 18C) as expressed in the inverse equivalence, $w(n) \sim R/n^x$, where x is close to 2 for large literary works such as *David Copperfield* by Charles Dickens, which contains a total of $T = 362892$ words. Simon showed that this relationship can be rewritten in terms of word *rank* (Figure 18A). Words with the highest occurrence are assigned the lowest rank. Thus the word with highest incidence is assigned a rank of one. The second most frequent word is assigned a rank of 2, and so on. The rewritten form of Zipf's law gives a measure of the rate and definiteness at which new context is generated in text. This formulation, known as Simon's model, gives a mathematical description of the underlying trend in Zipf plots by relating the total number of words, T , to the lexicon size, V (or number of different words appearing in the work) via various parameters, of which the exponent, ν , reveals the rate of context. The slope of a Zipf plot flattens out more towards higher word ranks (indicated by a straight line in Figure 18A) for literary works of large text length, T . Simon found a close relationship between the flattened slope of the plot, the rate at which new words appeared in a text and ν . He concluded that ν is implicitly linked to the rate at which context is generated in a given text. Higher values of ν implied a higher rate of context generation. Such works necessarily follow a less strict context.

The two models were recently shown to characterize the distribution of notes in four musical compositions (Zanette 2006), with decreasing values of context definiteness, ν , for works by Bach, Mozart, Debussy and Schoenberg. The result matches the chronological order and level of atonality of the four respective works, possibly a reflection of a change in attitude towards density of compositional con-

text in the shift from baroque through romanticism to atonalism. A surprising aspect of Zanette's finding is the implication that structural order in musical composition and in literary works can be revealed by the same quantitative measure.

Moving from literature and music to visual structure, I found that the same approach applies in measuring structural order of trees (Figures 18B & D). The trees depicted in Japanese scrolls and real garden trees manicured by Japanese gardeners all adhere to aesthetic ideals such as those outlined in Table 2. Hence, three questions naturally arise. First, the question of whether or not Zipf's law governs the population count of visual parts in such tree structures. Second, whether Simon's model would still provide an agreeable fit for tree data and thus estimate the definiteness of visual context in a tree structure. Third, whether this measure of context rate would reveal intrinsic structural differences between natural, inappropriately pruned, professionally pruned, and painted tree structures.

The Zipf-Simon approach can only be applied to trees once the lexical units of tree structure are defined. This is no easy task. Words and musical notes are denoted as distinct elements, whereas transitions from one branch to the next are ambiguous. The lexical units of a literary work is probably the clearest of all, although the Zipf count could conceivably also take word combinations, phrases and consecutively longer passages into account as units in a lexicon that relates to the larger scale arrangement of literary structure. In the case of musical compositions, the distinction between notes and musical phrases as individual lexical elements becomes more obscure. A single note seems to qualify as both a letter in the alphabet and as a word, to use an analogy with written language. Zanette solved the problem by denoting a unique lexical unit for each possible combination of pitch and duration of the musical notes appearing in a given composition, from which Zipf-Simon analysis was possible.

Following Zanette's line of reasoning, one would assign a visual lexicon of a tree based on distinct features characteristic of its most obvious subparts – branches. For example, the length, angle and average width of each branch should enable one to create a relatively large visual lexicon. But here two significant problems arise. First, as a tree matures, the constant flux of structure in all its parts give rise to limbs that do not consist of simple straight units obeying strict bifurcations. The branches on a real tree could be considered the equivalent of intricate paths carved by a history of events affecting the tree, such as availability of water and light, competition with other trees or branches, calamities such as pests, storms, cold and fire, and the endless structural shifts as all parts of the tree are minutely rearranged in the process of growing. In the undulating path of each branch one therefore has to decide whether the entire branch counts as one structural unit, or if each curvature maximum should denote unique branch components.

Bifurcations in trees complicate the decision. It could either signify two new structural units, or a side branch emerging from an ongoing parent branch. A tree

lexicon which could also take into account properties such as asymmetry, uniformity or self-similarity is even more evasive. On a more practical level, the difficulty in automating the construction of appropriate three dimensional data from biological trees prohibits the search for the most appropriate lexical units characteristic of tree structure.

Nevertheless, I show trial analyses using simple computer generated trees and trees depicted in classical scrolls. In each case, the lexicon was constituted by denoting each branch according to its length and angle relative to a zero degree vertical reference.

The inverse relationship between the number of branches, $w(n)$ which occur n times, and n , is clear when n and $w(n)$ are logarithmically plotted against branch

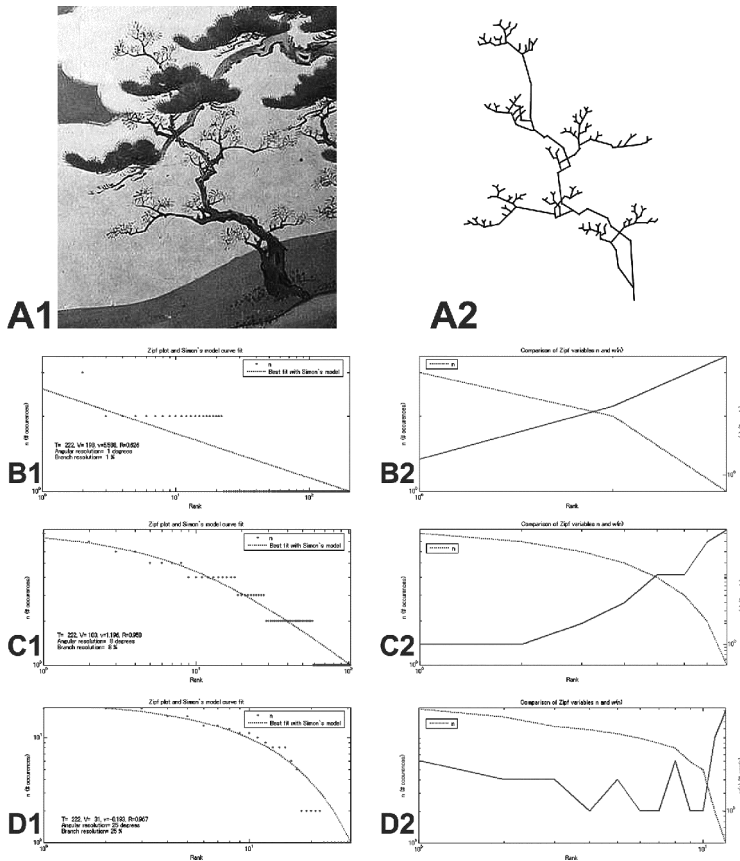


Figure 19. (A1) Illustration of a tree from the classical Japanese literary work, Tale of the Genji, and its line drawing (A2) with $T = 222$ limbs. While Simon's model provides better fits to data sampled at consecutively lower resolutions (B1, C1, D1 respectively), the inverse relationship between n and $w(n)$ in the Zipf plots (B2, C2, D2 respectively) deteriorates.

rank (Figure 18D). In the first example, the author intuitively redrawn a classical illustration of a tree (Figure 19A1) into a line diagram (Figure 19A2) from which discrete components and thus a lexical count was obtained. This tree clearly obeys Zipf's law (Figure 19B1) and is well predicted by Simon's model (Figure 19B2). But here an intrinsic difference between musical and literary lexicons and tree structure becomes apparent. While words, note pitch and note duration are given in a clearly discrete form, branch length and branch angle depends on resolution. The finer the resolution, the larger the lexicon, since a greater number of unique branches can be discriminated. As the resolution of angle and length decreases, the distinction between branches becomes less accurate and more branches with roughly similar orientation and length become grouped together under the same lexical unit. Hence V decreases. While T (total number of branches in the tree) remains constant, V , $w(n)$ and n vary. Zipf plots and Simon estimates are thus affected by the resolution at which visual structure is presented. Figures 19B1, C1 & D1 show how the Zipf plot of an illustrated tree, from the Genji scrolls, deteriorate for increasingly lower lexical resolution. Simon's plots of branch count, $w(n)$, against branch rank improve with lower resolution (Figures 19B1, C1 & D1). Simon's mathematical model provides a closer fit to ranked data as V decreases. But this also means that ν , or the definiteness of structural context, changes with resolution. Here ν decreases, or, context becomes stricter with lower resolution. In musical terms, the tree becomes less *atonal* with lower resolution. Actual Zipf plots (Figures 19B2, C2 & D2) deteriorate for increasingly lower lexical resolution.

The total number of units in the above example was small ($T = 222$). Nevertheless, the Zipf plots and Simon's curve fits remained acceptable over one order of magnitude of change in resolution. Further reductions in spatial resolution especially affect the variable $w(n)$.

A similar trend is observed with different simple computer generated trees (Figures 20~22) with comparable or larger number of branches than in the first example. Zipf plots deteriorate with lower resolution, while Simon's model provides better fits. This is a result of the fact that Simon's model better predicts trends in data sets for which $V \ll T$.

For a symmetrical tree with incremental changes in branch path length (Figure 20A) the inverse relation between $w(n)$ and n is more obscure, but the Zipf plots change surprisingly little over different resolutions, a result of the rigid compartmentalization of branching angles and the fact that $V \ll T$ from the outset.

Introducing length asymmetry at each branch node changes the appearance of the symmetrical tree and hence also its Zipf data points (Figure 21A). The inverse relation between $w(n)$ and n has improved somewhat from the symmetrical case.

Randomization of branch angles (Figure 22A) leads to a much larger variety of tree components, and hence a larger lexicon, V . In such a case, Simon's model should not fit the data as well as in the previous examples. Nevertheless, at highest

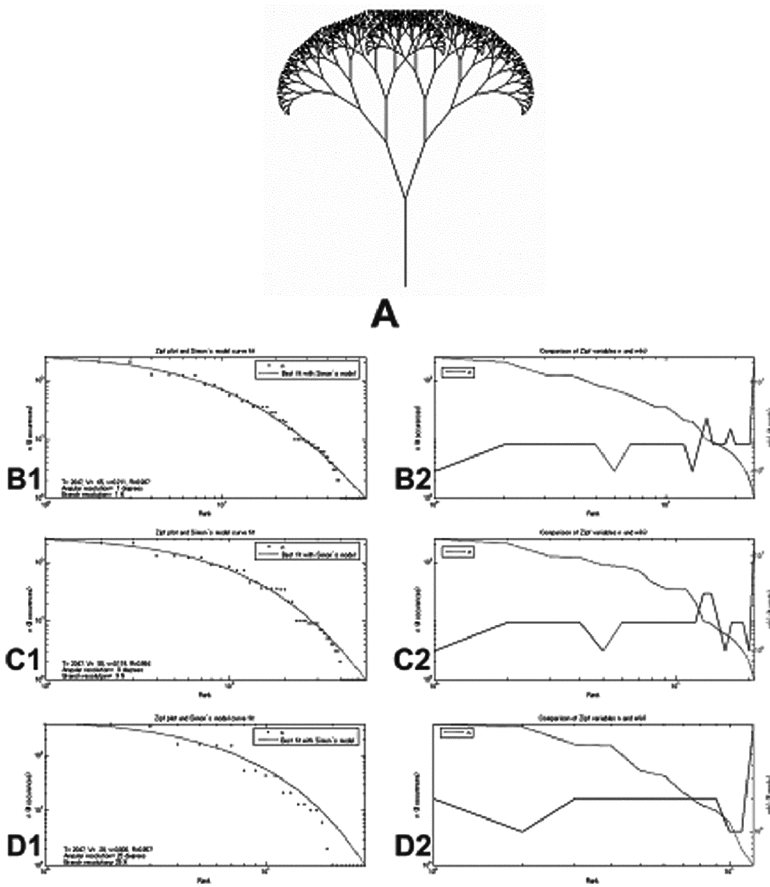


Figure 20. (A) Symmetrical computer generated tree with $T = 2047$ limbs. Simon's model fits to Zipf data points at consecutively lower resolutions (B1, C1, D1 respectively). The inverse relationship between n and $w(n)$ in the Zipf plots (B2, C2, D2) are not as clear as in the previous example.

spatial resolution, the variables n and $w(n)$ more closely approximates Zipf's law than in the previous cases with more ordered structures. At lower resolution the Zipf plots deteriorate as expected (Figures 22B2, C2 & D2).

The above examples demonstrate the influence of spatial resolution on Zipf plots and Simon's prediction of Zipf data points. This clearly sets apart the approach of lexical analysis in visual and non-visual structure. This difference further suggests that Zipf variables and Simon's models are most optimal at a given lexical resolution. The concepts of resolution and optimality of a lexicon, in this context, have not appeared where Zipf's law was applied to literary and musical compositions, since those formats are inherently independent of discretization. The

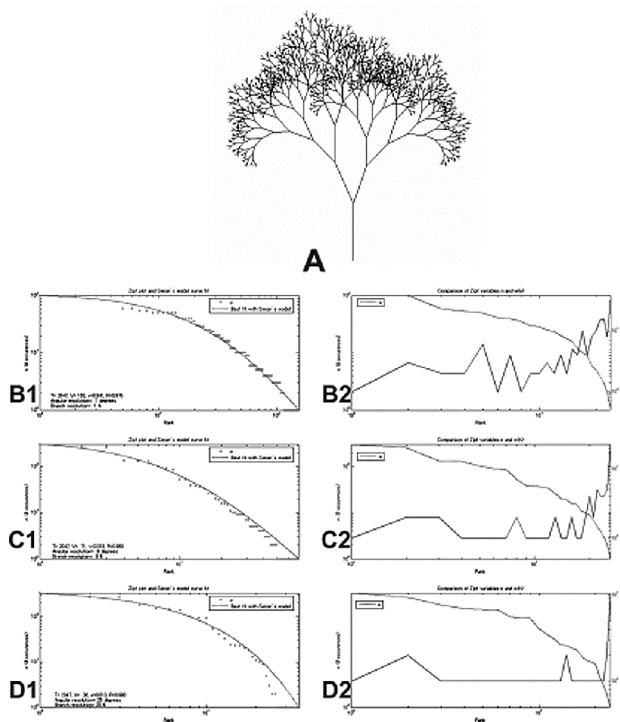


Figure 21. (A) Tree with asymmetrical branch length. Simon's model fits (B1, C1, D1) and Zipf plots (B2, C2, D2) are shown for different lexical resolutions.

value of ν for the optimal Zipf-Simon graphs can then be used in a more reliable comparison of the rate of context generation in different visual structures.

The last example (Figure 22) shows the optimal Simon's curve fittings and Zipf plots for eight trees, some of them already shown. For most of the examples shown here the optimal Zipf data points were obtained at a resolution of 4 degrees per branch angle and 4 percent of the maximal branch length. The highest rate of context generation among all these trees was measured for the classical illustration of a tree. Even trees with randomized branch angle paradigms had much lower context rates. This could reflect the fact that the illustrated tree used here was presented in a sparse canonical form.

Flat trimming of trees (Figures 23C1, 23H1) lowers contextual integrity (increased ν), while the purely symmetrical tree has the strictest context of all the trees. Its symmetrical trimmed version has lower contextual integrity, since flat trimming creates a larger diversity in the kinds of components from which the lexicon is compiled.

At first glance the Zipf plot data shown here resemble the inverse power law governing natural image statistics. The similarity originates from the fact that the

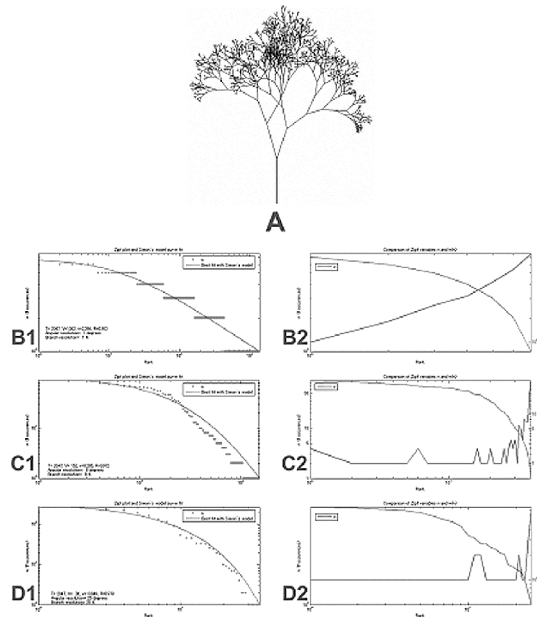


Figure 22. (A) Tree with asymmetrical branch length, and branch angle perturbed with uniform random noise. Simon's model fits (B1, C1, D1) and Zipf plots (B2, C2, D2) are shown for different lexical resolutions. A clear Zipf plot is obtained at highest resolution.

most frequently occurring structural tree components constitute a very small segment of the entire structural lexicon, whereas greater numbers of lexical categories occur only for a small number of times throughout a tree. The two trends differ significantly, however. While lowest spatial frequencies constitute the bulk of the image spectrum and thus contribute to the inverse power law of image statistics, any limb on a tree counts as one structural component, with equal energy regardless of how large or small that component is; the main tree trunk contributes as much as a single tiny twig. The fact that the occurrence, n , of a given tree branch type is inversely equivalent to the number of branch types, $w(n)$, that occur n times, relates from organization of the tree on an abstract level.

As applied here, the Zipf-Simon paradigm did not explicitly reveal differences between asymmetry, homogeneity and balance of visual structure, other than the above context ratings. Future work will consider how to make the approach more useful. For example, measurement of ν can be improved by more sophisticated testing of how closely both Zipf's law and Simon's model are obeyed. In essence, the results remain complex to interpret, since various factors play a role in the eventual structural analysis. However, the approach paves the way for direct comparison between tree structures (and visual structure, such as architectural facades

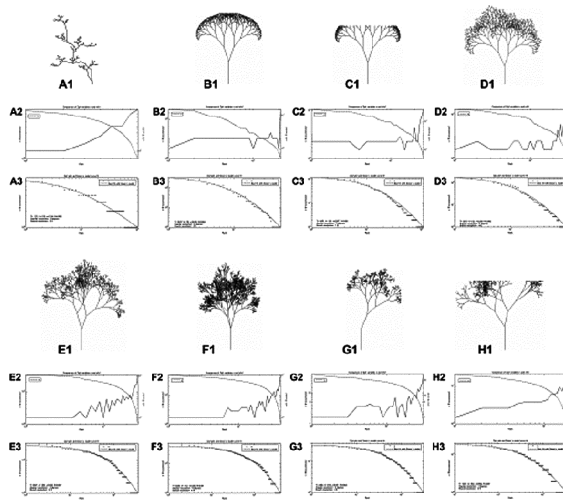


Figure 23. Optimal Zipf data points and Simon's model fittings for eight trees. (A) $\nu = 1.2$ for classical tree illustration. (B) $\nu = 0.21$ for symmetrical tree and (C) $\nu = 0.31$ for flat trimmed symmetrical tree. (D) $\nu = 0.3$ for asymmetrical branch length rule. (E) $\nu = 0.35$ for random dichotomous tree, (F) $\nu = 0.28$ for random trichotomous tree, and (G) $\nu = 0.33$ for random mixed mono/dichotomous branch rule. (H) $\nu = 0.6$ for trimmed random dichotomous tree.

etc.) and other compositions, such as literary works and music, and understanding the visual lexicon of a design.

A more appropriate lexicon of visual structure

As with literary and musical works, the full potential of Zipf-Simon analysis in visual structure will not be realized without a more comprehensive lexicon. Rather than only including single notes, words or branches, the lexicon should ideally also reflect relations between structural units. Literary and musical phrases, harmonic chords, and so forth, would qualify as higher level lexical entries.

In the previous section, branches were decomposed into straight line segments. Obviously, a more holistic level of lexical description is possible. The various forms of the red pine tree, as it appears in nature and in various guises in gardens (Figure 24), intuitively differ in their visual composition, and a more holistic approach could clarify these differences. For example, the sequence of structural deformations that a branch undergoes along its major axis would capture more holistic structural distinctions between branches. The types of hierarchical connectivity between branches could help one to define the holistic lexicon.

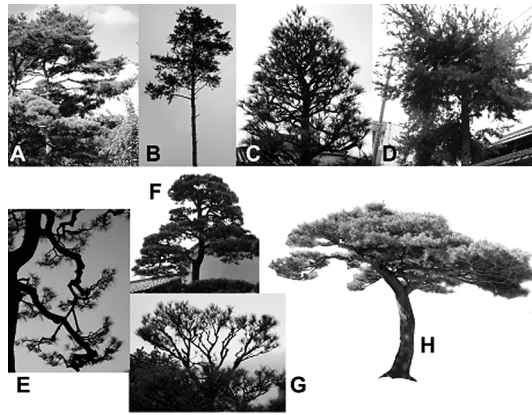


Figure 24. Wild red pines (A, B), and two examples of less fortunate pruning (C, D). Here, branches are too dense, inviting pests and structurally weakening the trees. The same red pine species is pruned along various styles (E, F, G, H), which can reach high levels of craftsmanship.

Although this level of analysis is still beyond the scope of this work, it would almost certainly be constructive in more clearly underpinning the structural differences between natural and artificially shaped trees, and in the latter case deepen our insight into how different styles of pruning affect the visual structure of a tree.

Conclusion

This chapter is intended as an overview of recent research on visual perception in Japanese dry rock gardens. For a more detailed understanding, the reader is invited to consult with earlier publications by Van Tonder et al. as cited in the text. However, in general I conclude that *karesansui* embodies a masterful design approach which brings the mind to complete the composition.

The collected classical teachings on Japanese garden design and their associated visual effects have been the inspiration behind this work. We have found substantial overlap between the visual features that classical gardeners manipulated and the perceptual grouping factors used by the Gestalt school to describe the process of figure-ground segmentation. Specifically, our investigation suggests that *karesansui* design simplifies the segmentation process, equalizes the salience of parts and wholes over a range of spatial scales, and balances the salience of figure and ground. It recruits mechanisms of perceptual completion to reconstruct visual features that are otherwise too complex to present in their totality, thereby achieving even more perceptually complex results with a sparse design composition. The abstract shape scaffold in such a design has many intentional

properties that support effortless and effective viewing and grasping of the visual complexity of the design. *Karesansui* design deliberately presents the viewer with generic views of parts and wholes. The discrepancy that arises in our natural ability to depict rectilinear structures and typical naturalistic compositions leads to a method of reconstructing maps of gardens, based on their illustrations in old texts, an important step when analyzing structural properties of garden designs that no longer exist.

Rocks and trees have served as the main poles of investigation, largely since the old texts treat them as the two fundamental components of the garden. Complexity in both elements is managed by using triangular units and odd-numbered visual junctions as building blocks from which the complex whole arises. Both are also often presented in half-occluded form, prompting the viewer to complete the shape according to their own internal wiring.

The two elements most significantly differ in their long term maintenance. Plants grow and change. Their shape results from a dynamic equilibrium maintained through the gardener wielding his mastery of pruning techniques. In doing so, the balance and spacing of a sparse group of branches guides the structuring of the tree, and literally becomes the way in which the gardener arranges clumps of leaves to produce the triangular units and appropriate visual junctions in this component of the garden design. Rock arrangement follows the opposite order, where rock clusters are manipulated in order to arrange the empty visual ground into a branching network similar to a carefully pruned tree.

Finally, I showed examples of formalization of tree structure in order to more deeply understand the impact of pruning on tree structure. Here, a direct link between literary works, music and visual structure is proposed, and various intuitions confirmed. A simple classical illustration of a tree displays a much higher degree of complexity than a relatively intricate computational tree with pseudo-randomized branches. Bilateral symmetry in trees significantly reduces their contextual complexity, while various degrees of randomization restore complexity. A sparse tree from a classical illustration nevertheless remains more complex: a surprising finding, given the canonical nature of the illustration. This approach paves the way for more detailed understanding of visual context, not only in pruned trees, but in visual design in general. Note that the use of Zipf's law and Simon's model are instructive in understanding the order and complexity of structures at an abstract level, but a richer description of shape is necessary if differences between the actual structures are to be intuitively understood.

Just as the definiteness of context decreases for more atonal musical compositions, one could expect an analogous situation in the garden. Specific forms of rock arrangement or tree pruning clarify the visual lexicon and allow the structure to stay closer to the theme or context. Inappropriate arrangements disrupt the visual lexicon and so corrupt the progression of context. More naturalistic gardens

would most likely find counterparts in the more atonal musical works, but I cannot imagine any natural scene approaching mere noise. For that the constraints of survival possibly do not tolerate much virtuosity.

Acknowledgements

The author thanks ikebana artist, Keiko Kishimoto, and the gardeners, Ryuuji Okuda, Hideki Okuda and Katsuyaki Ogawa for valuable discussions on arbo-real shaping. This research was supported by the Mitteleuropa Foundation and the Japan Society for the Promotion of Science (JSPS).

Note

1. For the sake of compactness, the terms 'garden' or *karesansui* will be used interchangeably to refer to Japanese dry rock gardens, unless otherwise specified.

References

- Akisato, R. (1799). *Miyako rinsen meishou zue*. 6 volumes. Kyoto.
- Akisato, R. (1828). *Tsukiyama teizouden gohen*. 2 volumes. Nagoya.
- Alexander, C. (1979). *The Timeless Way of Building*. Oxford: Oxford University Press.
- Alexander, C. (2002). *The Nature of Order: An Essay on the Art of Building and the Nature of the Universe*. Berkeley, CA: The Center for Environmental Structure.
- Arnheim, R. (1966). *Toward a Psychology of Art*. Berkeley: University of California Press.
- Arnheim, R. (1988). *The Power of the Centre: A Study of Composition in the Visual Arts*. Berkeley: University of California Press.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115–147.
- Blum, H. (1973). Biological shape and visual science (Part I). *Journal of Theoretical Biology*, 38, 205–287.
- Jiroh, T., & Keane, M. P. (2001). *Sakuteiki: Visions of the Japanese Garden – a Modern Translation of Japans Gardening Classic*. Tokyo: Tuttle Publishing.
- Kemp, M. (1990). *The Science of Art*. Hong Kong: Yale University Press.
- Kinchla, R. A., & Wolfe, J. M. (1979). The order of visual processing: 'Top-down,' 'bottom-up,' or 'middle-out.' *Perception & Psychophysics*, 25, 225–231.
- Kishi, F. (1994). The perspective system of Early *Uki-e*: Okumura Masanobus series of *O-uki-e*. *Journal of the Japanese Society for Aesthetics*, 6.
- Koenderink, J. J., & van Doorn, A. J. (1979). The internal representation of shape with respect to vision. *Biological Cybernetics*, 32, 211–216.
- Koffka, K. (1935). *Principles of Gestalt Psychology*. New York: Harcourt, Brace & Co.

- Kovács, I., Feher, A., & Julesz, B. (1998). Medial-point description of shape: A representation for action coding and its psychophysical correlates. *Vision Research*, 38 (15/16), 2323–2333.
- Kuitert, W. (2002). *Themes in the History of Japanese Garden Art*. Honolulu: University of Hawaii Press.
- Leyton, M. (1992). *Symmetry, Causality, Mind*. Cambridge, MA: MIT Press.
- Mandelbrot, B. B. (1977). *The Fractal Geometry of Nature*. New York: W. H. Freeman and company.
- Milner, A. D., & Goodale, M. A. (1995). *The Visual Brain in Action*. Oxford: Oxford University Press.
- Nakane, K. (1999). *Kyoto Meien Hyakusen*. Kyoto: Tankosha.
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9, 353–383.
- Nitschke, G. (1966). MA: The Japanese sense of place. *Architectural Design*, 36, 116–156.
- Nitschke, G. (1993). *Japanese Gardens: Right Angle and Natural Form*. Köln: Benedikt Taschen Verlag GmbH.
- Nitschke, G. (2000). *Japanese Gardens*. Köln: Benedikt Taschen Verlag GmbH.
- Palmer, S. E., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), *Attention and Performance*, 9 (pp. 135–151). Hillsdale, NJ: Erlbaum.
- Panofsky, E. (1927). Die Perspektive als 'symbolische Form' (Perspective as symbolic form, translated by C. S. Wood, 1991, New York: Zone Books). In *Vorträge der Bibliothek Warburg 1924–1925* (pp. 258–330). Leipzig & Berlin: B. G. Teubner.
- Shigemori, M. (1936–1939). *Nihon teien shi zukan* (Historical album of Japanese gardens). 26 volumes. Tokyo: Yuukousha.
- Shimoyama, S. (1976). *Translation of Sakuteiki: The Book of the Garden* (Tokyo: Town and City Planners). Attributed to Toshitsuna Tachibana, late 11th/early 12th Century.
- Shingen (1466). *Senzui Narabi ni yagyou no zu* (Illustrations for designing mountain, water and hillside field landscapes). Original teachings of the priest *Zouen* dating from the 11th century, transmitted by Shingen. Published in Sonkeikaku Library, Sonkeikaku Soukan series. Tokyo: Ikutoku Zaidan.
- Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, 42, 425–450.
- Tarr, M. J. (1995). Rotating objects to recognize them: A case study of the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review*, 2, 55–82.
- Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception & Performance*, 21, 1494–1505.
- Van Tonder, G. J. (2006, in press). Recovery of visual structure in illustrated Japanese gardens. *Pattern Recognition Letters Special Issue on Pattern Recognition in Cultural Heritage and in Medical Applications*.
- Van Tonder, G. J., & Ejima, Y. (2000). From image segmentation to anti-textons. *Perception*, 29, 1231–1247.
- Van Tonder, G. J., & Ejima, Y. (2003). Flexible Computation of Shape Symmetries. *IEEE Transactions on Systems, Man & Cybernetics, part B*, 33(3), 535–541.
- Van Tonder, G. J., & Lyons, M. J. (2005, in press). Visual perception in Japanese rock garden design. *Axiomathes Special Issue on Cognition and Design*. Elsevier Press.

- Van Tonder, G. J., Lyons, M. J., & Ejima, Y. (2002). Visual structure of a Japanese Zen garden. *Nature*, 419, 359–360. Manuscript supplement available at: http://www.ipc.kit.ac.jp/~gert/publications/selected_publications.html
- Watanabe, T., & Cavanagh, P. (1996). Texture laciness: The texture equivalent of transparency. *Perception*, 25, 293–303.
- Wertheimer, M. (1938). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A Sourcebook of Gestalt Psychology* (pp. 71–88). London, UK: Routledge and Kegan Paul.
- Willats, J. (1997). *Art and Representation: New Principles in the Analysis of Pictures*. New Jersey: Princeton University Press.
- Zanette, D. H. (2006). Zipf's law and the creation of musical context. *Musicae Scientiae*, 10 (1), 3–18.
- Zipf, G. K. (1949). *Human Behaviour and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley.

Thoughts on shape

Frederic Fol Leymarie

1. Introduction

What is shape?

Shape is the information, the structure induced by the presence of object elements in a space. An object manifests its presence by bits of data we refer to as “outline,” to be defined later. Such an outline can be obtained, for example, by a sampling process, such as when a laser camera scans the surface of a 3D object. Space itself is defined by its associated measuring sticks, or lack thereof, and the number of dimensions needed to enumerate its extent.

Shape is the “glue” between object and space, revealed by an interrogation of outlines. We interrogate an object by probing its outline. This inquisitive process we call a “transform” as it not only is a process which attempts to identify and characterise the object’s structure, but in doing so modifies, perturbs, the object and space itself.

Complexity in forms emerges by the grouping, assembly, concatenation of object elements. This complexity can be understood by defining a language for shape. Such a language could include atomic structures, the alphabet or a code, which when put together, via a grammar, create a word, and assembly of words create sentences. The alphabet and grammar can then be thought of as the essence of what the organisation of shape is.

In the following, I first define a few concepts to initiate our discussion by giving meanings to “space,” “object,” “outline,” “shape,” “form,” “structural field,” “shape probing,” “representation.” I proceed to explore some of the most popular transforms which can reveal the nature of shape and then specify classical geometries: the science of measure on space and its objects which leaves invariant certain properties under specific transforms. I then depart from the tradition and propose a dichotomy between two main classes of transform sequences, I call “horizontal” and “vertical,” which do not require invariance to specify and classify useful probes for shapes. While horizontal transform sequences refer to transforms

which operate along an outline, vertical ones act away from it. Along the way I provide a number of examples from the field of computer vision illustrating various transforms and their belonging to one class or the other. Finally I consider a few examples of “shape grammars” which permit to construct and analyse shape at varying degrees of complexity.

2. Preliminaries

Definition 2.1 (Space) *Entity having extent spanning a number of dimensions; i.e., such that we can refer to any constitutive locus – positions – of that entity. Space can be homogeneous, i.e., such that its dimension set is everywhere similar, or heterogeneous, i.e., such that its dimension set may vary locally, regionally. For example, consider the “classical” Euclidean space in 3D which is homogeneous, isotropic and of infinite extent, versus Einstein’s relativistic space which is heterogeneous as a function of mass (of objects).*

Definition 2.2 (Object) *A region of a given space, such as the Euclidean 3D space, E^3 , having some distinguishing characteristic. For example, a (homogeneous) region, x of E^3 , whose density, δ , is above some fiducial value, φ , i.e.: $\delta(x) > \varphi$; refer to the concept of solid shape and tolerances for shape by Koenderink (Koenderink 1990: Ch. 2).*

Definition 2.3 (Outline) (i) *Samples of an object delimiting its outer limits. For example, consider the lines (curves) marking the edges of a pattern in an image. Also, (ii) a sketch in which object features (not necessarily boundaries) are marked (without shading; e.g., a line-drawing of a figure).*

Definition 2.4 (Shape (in isolation)) *Description of the outline of an object.*

Such a “description” might simply consist in a quantitative study of the outline itself; e.g., the ordered enumeration of the contour points.

Alternatively, a “description” of some of the distinctive features extracted from the outline can define “shape;” e.g., the 3 corners of a triangle.

Thus, this description will depend on the method used to probe the object’s outline. As such, one may say that “shape is operationally defined” (Koenderink 1990: 15).

Definition 2.5 (Form) *Qualitative description of the shape of an object; e.g., the “round form” of the sphere or of an ellipsoid.*

Thus, “space” is a place-holder for content, while the notion of “object” corresponds to (i) the physical body (assumed) present in the scene, or (ii) the region of space under scrutiny. The notion of “shape” corresponds to the description of

the “available” outline of this object.¹ For example, it may be derived from some projection of the object in the scene onto the image (or retinal) plane. Finally the notion of “form” is a more abstract concept. Many shapes may have the same form. Shape is the “content” of form; i.e., it gives a more precise meaning or identity to a form. I think of form as a kind of “meta-shape,” a concept which is useful for object recognition.

Class of object outlines to be considered

In the following I will consider as “objects” sets made of points, curve segments, surface patches and volumetric elements.

N.B.: I will not define at this point, in a formal manner, what I mean by a curve or a surface or a volumetric object. I shall not commit myself (yet) to a particular method, such as in differential geometry, where these objects are defined through parametrisation under compatible diffeomorphisms (i.e., coordinate charts). Such methods of definition tend to impose strong constraints on the way one looks at the objects being scrutinised. Along the way, however, I will introduce and use some of the terminology of differential geometry, e.g., to express relevant concepts such as tangents and vector fields.

Shape can also be defined within a context or environment. For example, shape can be extended to embody the influence of “gravity,” i.e., the orientation of an object with respect to an horizon or reference frame, and its relative position with respect to that horizon. Consider for example a “diamond” shape which is distinguished from a square or parallelogram only by its orientation with respect to an horizon or ground plane. This is a famous “phenomenon” first identified by Ernst Mach in the late 19th century and later “refined” by Goldmeier in the 1930’s (Leyton 2001: Fig. 8.3). Refer also to the related notions of “visual weight” and “visual direction” from the Gestalt theory of perception: a measure of perceptual forces exercised by an object on and via its visual context (Arnheim 1974).

Definition 2.6 (Shape (within a context)) *Description of the outline of an object under the influence of its environment, or of other objects.*

The study of the shape of an object in isolation or within a context, a scene, can be unified via the concept of a “structural field.”

2.1 Shape as the structural field of an object

Concept 2.1 (Shape as a structural field) *Shape is the structure of the field existing “around” an object. This field may consist of geometric entities – such as curvatures, singularities (of some appropriate mappings), gauge figures (e.g., Dupin’s indica-*

trix) – which may exist in association with each sample of the object being scrutinised. In order to identify or measure such entities, we will have to “probe” the field.

Koenderink’s point of view

Paraphrasing Jan Koenderink (Koenderink 1990:§2.2):

One can say that objects *induce a* (geometrical) *field* around them which can be *probed operationally*. The *structure of this field* is the *shape*. Shape is a property of *both* the object and the probing method. Furthermore, one can say that shape will depend on the *perception*, i.e., the mode of interaction (probing) and the expectations (models, theories).

Concept 2.2 (Probing shape and the uncertainty principle) *Since “shape” will also depend on the probing method, and not simply on the object’s structure, we have here an example of a variation on the “uncertainty principle” of Heisenberg in action. We will be as precise about our identification and measure of the structural field of an object as the selected probing method permits to.*

We will have to seek for probing methods which, although they may modify (and not just identify or measure) the object’s field, preserve as much as possible this structure, or emphasise certain of its features. Also, we will have to identify which structures are most relevant for shape analysis. This selection can only be performed by considering jointly the effects of probing the field. Note also that, if we want to take into account the context for an object, say its orientation relative to a ground plane, the probing method will need to be defined to also consider this influence. Various probing methods will emphasise various features in the form of a “representation.”

Definition 2.7 ((Shape) Representation) *Mathematical abstraction of an object making explicit special selected features.*

Typically, a representation results from applying a probing method to the outline of an object, revealing certain shape features, such as the corners of a polygon. Note that we can further consider the representation of a representation and build a hierarchy of structures for an object, i.e., make its shape available at various levels of abstraction. For example, consider a “donut” object in Euclidean space (a torus). An initial representation may capture the closed tubular surface structure from original surface samples. This representation can be further probed to reveal the central circular axis of the tore-like object and the associated hole it creates in space.

A final refinement in my definition of “shape” is necessary to capture the notion of complexity, such as due to the concatenation of parts or due to deformations and growths.

2.2 Shape as the recoverable structural field of an object

Shape as a recoverable structural field

Paraphrasing Michael Leyton (Leyton 2001:§2.23, 75):

Shape is the recoverable “structural field” of a machine whose initial state space is the object’s outline.²

A “machine” is a pair of functions acting on a given input set (of parameters, data) (Leyton 2001:§1.15, 30): such that (i) one function specifies the states of the machine being transformed, and (ii) the other function specifies a new output set produced. The concept of a “machine” emphasises that “shape” can be cast as a *dynamic phenomenon*.³

In the following section, important examples of probes used to reveal the shape of an object are first considered. Such probes are called “shape transforms” or simply “transforms.” After some preliminary definitions, a number of important (practical) properties for shape transforms are considered. This preliminary study will set the stage to consider sequences of shape transforms and construct a bi-partite classification (in §4) inspired by applications in computer vision used to reveal the shape of objects.

3. Transforms for shape

Definition 3.1 (Domain and range) *Let us call the space (or region of space) containing the outline or representation being probed, the domain, while the space (region) containing the resulting outline or representation is called range. Let us denote the former by D and the latter by R .*

Concept 3.1 (Shape transform) *A mapping F from D to R with properties to be specified latter (§ 3.1). Thus,*

$$F : D \rightarrow R.$$

Concept 3.2 (Shape operator) *The mathematical (operational) object, ρ , which permits to implement the transform F . Thus,*

$$F : D \xrightarrow{\rho} R.$$

For example, the operator could be a convolution kernel in the context of a smoothing transform.

Remark 3.1 (Spaces for domain and range) *The domain and range of a shape transform may lie in the same space or not. E.g., in the case of a “projection,” the range lies in a subspace with respect to the original space of the domain.*

Note, however, that in many practical situations, both the domain and range will lie in the same space, i.e., the transform is an instance of a *general deformation*: 1-to-1 mappings of a space onto itself.

Question 3.1 (What is the purpose of a transform?) *Typically, one uses the concept of a transform to simplify, or break down, or map to a more explicit representation, the problem being attacked. A simplification may have different meanings depending what are the goals to be achieved. A transform may make more explicit some property(ies) of the shape (or the domain); e.g., it could amplify the presence of linearities of a planar object such as a square; a classical example here is the Hough transform.⁴ Alternatively, a transform may look more like a filter which removes some undesirable properties; e.g., smoothing a surface with convolution kernels, this to attenuate noise or to interpolate data. A classical example here is the Fourier transform which first maps a signal or image from the spatial domain to the frequency range where certain spatial frequencies can be selected; e.g., in the context of low, high or band pass filtering.*

Remark 3.2 (Shape interrogation in CAD-CAM) *In the field of CAD-CAM,⁵ the term “interrogation” is used for an equivalent notion of shape transform (Patrikalakis & Maekawa 2002: Ch. 8): “interrogation is the process of extraction of information from a geometric model,” where the latter corresponds to the notion of outline.*

3.1 Examples of shape transforms

In the following I enumerate important shape transforms in terms of their distinguishing properties.

3.1.1 Invertible transforms

It is often argued that a transform for shape should be invertible, i.e., the mapping F should have a unique inverse $F^{-1} : R \rightarrow D$, such that $F^{-1} \circ F = I$, where I denotes the identity, and $\circ$ the composition of functions. Why so? Because it is believed that if no information is lost (since we can then always recover the original data) the transform F should be a powerful one in the sense that it is *information-preserving*.

Involutive transform. An *involutive transform* is a special case of an invertible transform, such that it is its own inverse. Apply it twice and you are back to the original object, i.e., $F^{-1} \equiv F$, thus $F \circ F = I$.

A famous example of such a transform is the *Legendre transform* which maps functions on a vector space to functions on the dual space.

Definition 3.2 (Legendre tr. for functions of 1 variable) For a convex curve, $y = f(x)$, s.t., $f''(x) > 0$, consider the transformation of a curve from point coordinates $(x, f(x))$ to (tangent) line coordinates $(p, g(p))$.

Example 3.1 (Legendre tr. of polygonal contours) Let $f(x)$ be a convex polygon. Consider the graphs of f and g , i.e., the traces of f and g in the 2D spaces (x, f) and (p, g) . Then $g(p)$ is also a convex polygon, in which the vertices (edges) of $f(x)$ correspond to edges (vertices) of $g(p)$.⁶ Thus a corner is transformed into a segment (e.g., see Arnold 1989: Fig. 44, p. 62).

Definition 3.3 (Legendre tr. for functions of 2 variables) For a surface parameterised by a Monge patch, $u = u(x, y)$, which is non-developable (i.e., its Gaussian curvature is nowhere vanishing, and thus it cannot be spread flat on a plane), consider the transformation of a surface from point coordinates $(x, y, u(x, y))$ to (tangent) plane coordinates $(\xi, \eta, \omega(\xi, \eta))$.

Remark 3.3 The Legendre Transform is different from a mere coordinate transform, since, rather than assigning to a single point another point, it assigns to every surface element (x, y, u, u_x, u_y) a surface element $(\xi, \eta, \omega, \omega_\xi, \omega_\eta)$.

Remark 3.4 (From Lagrange's to Hamilton's mechanics) It is by means of a Legendre Transform, that a Lagrangian system of n 2nd order differential equations is converted into a remarkably symmetrical system of $2n$ 1st order differential equations, the "Hamiltonian system of equations" (or "canonical equations").

Note that, with respect to the previous comment on the "purpose of a shape transform," it appears that there is potential for some conflict of interest, when considering that a "good" transform should be *invertible*. For example, if one desires to "simplify" the shape representation (and eventually, description) it would come as no surprise that some information or level of detail is lost in the process. It might even be desirable to eliminate (or modify) some data which would be "irrelevant" for further analysis in any case; e.g., details existing at "too" fine resolution to be perceptually significant.⁷

3.1.2 Idempotent transforms

An idempotent transform is such that if you apply it more than once and sequentially (i.e., the output of one transform becomes the input for the following transform) you get the identity: nothing happens to the data anymore.

In other words, the composition of the transform F with itself yields F again (and forever):

$$F \circ F \equiv F,$$

or, in terms of the corresponding operator, ρ ,

$$\rho^2 = \rho.$$

Remark 3.5 (Fixed points of idempotent transforms) *Note that, all the points in the range of an idempotent transform, i.e. $F(\mathbf{x}) = \mathbf{y} \in R$, are fixed points.*

Jean Lagarde put forward the following informal argument in favour of idempotency (Lagarde 1990: Ch. 2):

If the goal of the operator is to recover a valid interpretation from initial data, then applying the process one more time on the supposedly valid interpretation should not change it.

A few examples of idempotent transforms are listed below.

Projection. The *projection*, $F \equiv \pi$, which maps a set $U \in D$ onto a subspace R^* of R , i.e., $\pi : \rightarrow R^* \subset R$ is idempotent. E.g., in Euclidean 3-D space, think of the orthogonal projection onto the x - y plane: $\pi : (x, y, z) \in E^3 \rightarrow (x, y) \in E^2 \subset E^3$.

Linearisation. A transform, $F \equiv l$, used in linear algebra of *vector spaces*, which is such that $\mathbf{v} \in R(l) \Rightarrow l(\mathbf{v}) = \mathbf{v}$. Thus, the corresponding matrix (operator), $\rho \equiv A$, is such that $A^2 = A$ (Bamberg & Sternberg 1988:38).

This linear operation is *non-singular* only if $A = I$, i.e., the identity, in which case l is a trivial mapping. Otherwise, it is *singular*, i.e., such that $\text{Det}(A) = 0$. E.g., if l is such that $l : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ and A is singular but not the zero matrix, then l maps the plane, $\mathbb{R}^2$ onto a line. If $A = 0$, then l maps the whole plane onto the origin.

An important application is the decomposition in a *Linearly Independent* (LI) set, i.e., the projection on basis sets/functions. E.g., think of an orthonormal representation such as the Fourier decomposition.

Mathematical morphology transforms. The *closing* and *opening* operators of mathematical morphology are such that they fill-in or remove details of size smaller or equal than a structural element B (Serra 1982; Soille 2004), which can be seen as a “probe” or tool to sculpt the original dataset.⁸ Applying these operators again does not further change the data which is said to have been *regularised* – up to the scale and layout defined by B .

Edge detection. Edge detection can be postulated as an idempotent transform (Owens et al. 1989; Ronse 1993). The transform F acts on the space of gray level

images, $D \equiv \mathcal{I}$, and has for its range a *binary feature map* corresponding to the edges and lines originally in $I(x, y) \subset \mathcal{I}$. Note that $R \equiv F(I)$ is also a valid image, i.e., $F(I) \subset \mathcal{I}$. However, applying F once more should not change the result: $(F \circ F)(I) = F(I)$.⁹

Remark 3.6 (Optimisation) Local energy minimisation *can be seen as an idempotent transform, a local minimum being a fixed point of the cost function (operator)* (Lagarde 1990: 15). *Applying the transform, once a local minimum has been reached, will not change the result.*¹⁰

Essentially, any transform implying an iterative and optimising process, which is defined as having to *converge* on some fixed point, can be casted as an idempotent transform.

3.1.3 Separable transforms

Transform in a coordinate system adapted to the boundaries¹¹ of the shape being transformed or to the domain acted upon. E.g., a transform acting on the real plane, $\mathbb{R}^2$, being such that:

$$F(x, y) = F(x) \circ F(y) = F(y) \circ F(x) .$$

Convolution kernels used in image processing often are selected because they share this property. E.g., smoothing and gradient filters based on the Gaussian kernel are more efficiently performed by using the separability of the Gaussian function (Burt 1981, 1983).

3.1.4 Linear transforms

A transform such that

$$F(ax + by) = aF(x) + bF(y) , \quad a, b \in \mathbb{R} .$$

A linear transform permits one to use the *principle of superposition*, i.e., arbitrary linear combinations of data can be transformed separately or altogether.

Note that a linear transform always possess a *matrix representation*, A , which can be taken as the operator. In particular, the composition of two linear transforms corresponds to matrix multiplication.

Affine transform. A linear transform *without* any preferred origin. Any affine transformation can be written as a (non-singular) linear transform followed by a translation.

An affine transform does *not* preserve distances or angles, but it does preserve *parallelism* and *ratios of lengths* (i.e., the ordering of points).

Conformal transform. A linear transform which preserves *angles* and *orientation* (i.e., such that $\text{Det}(A) > 0$), but not necessarily distances (Bamberg & Sternberg 1988: §2.1).¹²

Orthogonal transform. A linear transform which preserves *distances* (and, as a consequence, *angles* as well). It also preserves “ordinary” *scalar products*, $(\bullet, \bullet)$, i.e., such that they obey the following three properties: (a) *bilinearity*, (b) *symmetry*, and (c) *positive definiteness*.¹³ E.g., rotations, translations (and thus rigid motions), and reflections through the origin are orthogonal transformations. An *Euclidean transform*, i.e., equipped with the Euclidean metric and scalar product, is an example of an orthogonal transform.

Lorentz transform. A linear transform which preserves a *quadratic form*, Q , whose scalar product $Q(\bullet) = \langle \bullet, \bullet \rangle$ is symmetric, bilinear but *not necessarily positive-definite*. Associated with Q is a *symmetric matrix*, A , i.e., such that $A = A^T$, and thus $Q(\mathbf{v}) = (A\mathbf{v}, \mathbf{v}) = (\mathbf{v}, A\mathbf{v})$.

In the context of *special relativity*, the Lorentz transform must carry the *light cone* (or *null cone*), $\{\mathbf{v} : Q(\mathbf{v}) = 0\}$, onto itself. The corresponding affine transform is called the *Poincaré transform* and the corresponding metric is called the *Minkowski distance*, where the Euclidean circles are replaced by hyperbolas (Bamberg & Sternberg 1988: 154–157).

The Lorentz transform is typically used to relate the space and time coordinates (or in general any four-vector) of one inertial reference frame, A , into those of another one, A' , travelling at a relative speed of v to A along some axis, say the x -axis. This is the natural set-up to study the position, brightness and shape of images in arbitrary general relativistic spacetime, e.g., when considering gravitational lensing, i.e., how massive objects like stars bend light (Perlick 2004).

Symplectic transform. A linear transform which preserves the *symplectic scalar product*, $\omega(\bullet, \bullet)$, a bilinear, not necessarily positive-definite, and *anti-symmetric* (or *skew-symmetric*) product; thus, $\omega(\mathbf{v}, \mathbf{w}) = -\omega(\mathbf{w}, \mathbf{v})$. Note that ω is a 2-form, i.e., it takes its (real) values on pairs of vectors, and is a *skew-symmetric* operator. In (differential) geometry, it represents the *oriented area* of the parallelogram spanned by the two vectors it acts upon. In particular, in $\mathbb{R}^3$, it corresponds to the *vector product* of vector analysis.

Symplectic transforms are useful in geometric optics and Hamiltonian mechanics, and, in particular, are relevant to wavefront (or grassfire) propagations where one propagates fronts, i.e., stacks of oriented one-forms, while preserving their areas.¹⁴

3.1.5 Differential transforms

Transform which maps the outline to a differential representation. For example, the most common such transform maps an outline to its tangent space, where the best first order linear representation is obtained at each locus of the outline.

Lie group transforms and Hamiltonian flows. Lie group transforms are linear differential ones which obey a particular constraint, i.e., the Jacobi identity. Essentially, the latter states that flows (associated with tangent vector fields) *commute*, i.e., loosely speaking, if one travels along two flows, L_1 followed by L_2 , on a manifold, M , one ends at the same point than if one had travelled first along L_2 and then along L_1 . More on this topic and its relation to vision problems can be found in the works of William C. Hoffman (Hoffman 1966, 1978, 1984) and his followers (Ferraro & Caelli 1988; Papathomas & Julesz 1989; Hansen 1992).

Furthermore, certain classes of global deformations (of a continuum) can be expressed in terms of Lie derivatives, i.e., certain linear operators on vector fields (associated to flows) which indicate how differential forms change along integral curves on these vector fields. Lie transforms are instances of “horizontal transform sequences,” as we shall see later (§4.1).

Interrogation methods in CAD-CAM. Shape interrogation methods in CAD-CAM can be classified by the “order of the derivatives of the curve or surface position vector” involved (Patrikalakis & Maekawa 2002: Ch. 8). “Ray tracing” is an example of a first order method used to compute a shaded image of a surface assuming the reflected light is a simple (linear) function of the angle θ between the surface normal and the light direction vector (from a point light source). Examples of second order methods include “curvature plots” and “inflection point” maps for the interrogation of planar curves, “surface curvature” maps and “geodesics” (curves of zero geodesic curvature) for the interrogation of surfaces. Third order methods include “torsion” maps of space curves, and “stationary points (extrema) of curvature” of curves and surfaces. “Stationary points of the torsion of space curves” is an example of a fourth order method.

3.1.6 General deformation of a continuum

A non-necessarily-linear, non-necessarily-affine transform, where an object, O , is deformed via a mapping, F , that carries each material point, $\mathbf{p} \in O$, into a point $F(\mathbf{p}) = \mathbf{x} \in R$. This is a natural set-up for kinematics in the context of *continuum mechanics*.¹⁵ The shape transform, F , is called a *deformation*, and in general it is assumed one-to-one and smooth (thus, differentiable). An important special case is the *homogeneous* deformation, where the *deformation gradient*, ∇F , is constant: thus, global and local motions are the same. It can be shown that any

homogeneous deformation, F_{homo} , can be uniquely decomposed into a *rotation* about a fixed point, G , and a *stretch*, S_r or S_l :

$$F_{\text{homo}} = G \circ S_r = S_l \circ G.$$

Note that, in the special case of a *rigid* (and affine) transform/deformation, the gradient of F_{homo} reduces to a (constant) rotation with matrix representation $\mathfrak{R}$, i.e., $\nabla F_{\text{homo}} = \nabla G \equiv \mathfrak{R}$.

There are three main concepts used in the elementary study of general deformations: forces (operators), conservations laws (constraints), and constitutive assumptions (simplifications).

– **Forces** A force, $\mathbf{f}$, in this context can be equated with our notion of the operator of the transform; i.e., $\rho \equiv \mathbf{f}$. Three types of forces are considered in continuum mechanics:

1. Contact force between separate parts of the object, O , denoted $\mathbf{f}_c$.
2. Contact force exerted by the environment on the boundary of O , denoted $\mathbf{f}_e$.
3. Internal force exerted on the interior points of O , denoted $\mathbf{f}_i$.

The forces $\mathbf{f}_c$ and $\mathbf{f}_e$ acting on the boundary of O , i.e., on the surface ∂O , are represented through the use of a spatial tensor field, $\mathbf{T}$, called *stress* (or Cauchy stress). Note that, forces like “pressure,” “tension” or “shear” are all special instances of stresses.

In computer vision a simple model for deformable curves and surfaces under tension, called “snakes,” has been used to represent outlines and track them in dynamic imagery such as video sequences (Kass et al. 1988; Terzopoulos et al. 1987; Leymarie 1990; Leymarie & Levine 1993), as well as to simulate wave propagation constrained by a potential field (Leymarie & Levine 1992b).

– **Conservation laws** Various conservation laws are defined within the realm of continuous mechanics. These are general laws which are usually helpful in imposing constraints on the deformation F . Examples of such laws – which are all, really, different instances of the same fundamental conservation principle – are:

- Local and global *conservation of mass*.
- *Conservation of energy* (i.e., of the sum of potential and kinetic energies).
- *Conservation of momentum*, linear and angular, also called *momentum balance laws*.
- *Conservation of power* (i.e., the power expended equals the sum of stress power and rate of change of kinetic energy).

– **Constitutive assumptions** Conservation laws do not distinguish between *different types of materials*. To distinguish different types of material behaviour,

additional hypotheses, called *constitutive assumptions* are made. These are of three main types:

1. Constraints on the *possible deformations* the body may undergo. E.g., an isochoric motion (i.e., such that the object's volume remains constant).
2. Assumptions on the *form of the stress tensor*, T . E.g., a pure pressure or a pure tension, like in the case of the snake (or “active contour”) model of computer vision (Kass et al. 1988; Terzopoulos et al. 1987; Leymarie & Levine 1993).
3. Constitutive equations *relating the stress to the deformation in time*, i.e., to the motion; e.g., giving the pressure as a function of density as in the case of an *elastic fluid*.

3.1.7 Wavefront propagation and shock transforms

Imagine another type of general deformation where *no 1-to-1 mapping restriction is imposed*, i.e., many-to-1 or 1-to-many mappings are allowed or even desirable. E.g., there will be loci in the range, R , where points in the domain, D , are mapped to singletons – we say a *shock* occurred. In other words, *no continuum restriction* is put on the material being deformed.

Continuous wave propagation in geometrical optics. Consider a point $\mathbf{q}_0$ in Euclidean space E^3 at time t_0 , taken to be the generator of a disturbance transmitting itself “locally” (Arnold 1989:250). At time $t_0 + \Delta t$ the disturbance is propagated through space in the form of a surface, which we denote by $\varphi(x, y, z, t) = L$, representing the **wavefront** for each time t , where L is some constant. At any later time, the new wavefront can be obtained as a collection of disturbances from a previous wavefront; this is the geometrical construction first discovered by Christiaan Huygens in 1678.

Lemma 3.1 (Huygens’ principle) *Let $\varphi_{\mathbf{q}_0}(t)$ be the wavefront of the point after time t . For every point $\mathbf{q}$ of this front, consider the wavefront after time t_1 : $\varphi_{\mathbf{q}}(t_1)$. Then, the wavefront of the point $\mathbf{q}_0$ after time $t + t_1$: $\varphi_{\mathbf{q}_0}(t + t_1)$, will be the envelope¹⁶ of the fronts $\varphi_{\mathbf{q}}(t_1)$ such that $\mathbf{q} \in \varphi_{\mathbf{q}_0}(t)$.*

The wavefront is thus the boundary of the set of all points $\mathbf{q}$ to which “information” from a given source can travel in time less than or equal to t (Arnold 1989:249).

Huygens’ metaphor can be transferred to the discrete domain without difficulty, e.g. by considering the Minkowski sum of small discrete spherical sets with a discrete front at time t . This is precisely the realm of 3D Mathematical Morphology, where the spherical sets are called *structural elements* (Serra 1982). The problem is one of inefficiency: Minkowski sums with isotropic structural ele-

ments – required to ensure Euclidean results – have too much overlap. Consider then the alternative viewpoint of Pierre de Fermat’s principle (circa 1660), which states that:

Lemma 3.2 (Fermat’s principle) *Information travels along rays from a point $\mathbf{q}_0$ to a point $\mathbf{q}$ in the shortest possible time.*

Rays are such that their gradient vector is normal to the wavefront, i.e., the direction of the ray $\mathbf{p} = \nabla S$, where S is the (optical) “path length” from a source $\mathbf{q}_0$ to $\mathbf{q}$, a function such that $S_{\mathbf{q}_0}(\mathbf{q}) = t$. The (level) set of all such ray paths thus coincide with the wavefront $\phi_{\mathbf{q}_0}(t)$. Rays and wavefronts are related precisely through the refractive index v , which provides the (local) metric structure of the medium: $d\phi \cdot \nabla S = v$ (Stavroutidis 1972).¹⁷

Eikonal equation of geometric optics. Consider the special case where the index is taken to be isotropic and homogeneous. Then, the direction of motion of the rays and the wavefront coincide (Arnold 1989); thus

$$d\phi \cdot \nabla S = \|d\phi\| \|\nabla S\| = 1$$

Note that, considering a (light) ray travelling at constant speed c , since the index is fixed, we also have that $\|p\| = \|\nabla S\| = c$, leading to the classical “eikonal” equation for the wavefront (Stavroutidis 1972):

$$\|d\phi\|^2 = \left(\frac{\partial \phi}{\partial x}\right)^2 + \left(\frac{\partial \phi}{\partial y}\right)^2 + \left(\frac{\partial \phi}{\partial z}\right)^2 = 1/c^2 \quad (1)$$

Without loss of generality, we can take the constant $c = 1$, that is, we propagate wavefronts along straight rays with *unit* (relative) speed. Thus, the **normal field** to any wavefront is identically ∇S and each wavefront moves along this field by a unit distance step for each time step. We can therefore *identify time with distance* for an accurate simulation of wavefront propagation. Given an initial set of wavefronts or set of generators $\mathbf{Q} = \{\mathbf{q}_0 : \phi_{\mathbf{q}_0} = 0\}$ we can simulate wave propagation, according to the eikonal equation (1), by computing *point-to-set* (Euclidean) distances:

$$L(\mathbf{q}, \mathbf{Q}) = \min_{\mathbf{r} \in \mathbf{Q}} \|\mathbf{q} - \mathbf{r}\| \quad (2)$$

Note that, we may consider the set of all wavefronts, Φ , from an initial time (distance) $t = 0$ to a final time of propagation $t = T \equiv L_{\max}$, where $L_{\max}$ is the maximal distance travelled by the wavefront. The set Φ then fills-in a portion of space according to the distance function (2).

We call this volume of space filled with distance values by wave propagation, “Distance Hypersurface,” following Goodman who introduced the term in 2D (Goodman 1964).¹⁸ This is equivalent to an embedding of the wavefronts in a 4D space, where the fourth axis is taken for time or distance. The Distance Hyper-

surface is the graph of $\tau = \Phi(x, y, z)$, τ being taken along the time (or distance) axis.

Definition 3.4 (Regular distance hypersurface) *A hypersurface, $\tau = \Phi(x, y, z)$, is called a Regular Distance Hypersurface over a region $\mathfrak{R}$ of 3D space E^3 , if at each point of $\mathfrak{R}$ it is a “local distance hypersurface.” An hypersurface is a local distance hypersurface at a point $\mathbf{q} \in E^3$ if there is:*

1. *a spherical neighbourhood $\mathfrak{N}$ of $\mathbf{q}$,*
2. *a smooth surface $\mathfrak{S}$ which divides $\mathfrak{N}$ into two sub-regions, and*
3. *a constant l ,*

such that $L = \Phi(x, y, z) - l$ is the directed (or signed) distance from $\mathfrak{S}$ for each point in $\mathfrak{N}$.

This graph takes single values almost everywhere, according to equation (2), except at those points where wavefronts meet. This is where regularity is lost and singularities in the distance mapping occur. The latter define the symmetry points constitutive of a kind of 3D skeleton or symmetry set (§4.2). If we think in terms of the physical metaphor of light rays, this is where such rays meet to define the “burning points” or *caustics* of Archimedes.¹⁹

Another way to interpret the above concept is via the Implicit Function Theorem. If the gradient $\nabla\Phi \neq 0$ at some point $\mathbf{q}$, then it implies the regularity of Φ in a neighbourhood of $\mathbf{q}$. This only requires that Φ is one time continuously differentiable on the respective open set. Equivalently, having a Lipschitz continuous gradient ensures that the “symmetry points” stay away from the distance hypersurface for some distance (Wolter 1985, 1992).²⁰

The process of computing such a distance hyper-surface for a discrete dataset, is commonly called “distance transform” or “distance mapping” in image processing. When using the Euclidean metric, such as in equation (2) above, it is called “Euclidean Distance Transform” (EDT) (Danielsson 1980; Leymarie & Levine 1992a; Leymarie & Kimia 2000).

Shock transforms. As a somewhat concrete example, consider an outline taken as an initial wavefront. Then, take the transform which moves this wavefront according to the eikonal equation (1). This is a continuous deformation, also called an *evolution* for the (wave)front. There will be times and places where material points of the wavefronts collapse on, i.e., where a front self-intersects or becomes singular (non-smooth), or where two or more wavefronts meet each other. In terms of a wavefront propagation, this is called a “shock.”

A hierarchy of shocks of different types can be defined depending on the nature of this “collapse” or coalescence of material points. For example, in the case of shocks occurring in the context of the *planar curve evolution* problem, refer to

the early works of Blum (Blum 1967, 1973), Montanari (Montanari 1969), and more recent works of Kimia et al. (Kimia 1991; Siddiqi & Kimia 1996; Siddiqi et al. 1999a, b; Kimia 2003). Shock transforms are instances of “vertical transform sequences,” as we shall see later (§4.2).

Having described a number of commonly used shape transforms, I will now consider a classical way of *classifying* them via “geometries.”

3.2 Geometries for shape: Invariance under transforms

I follow here the classical definition of a geometry as the *study* of those *properties* of subsets of a region, D , which are *invariant* under the application of some specified transform. For example, one can consider *affine*, *Euclidean*, *projective*, or *symplectic* geometries for shape.

Felix Klein’s Erlangen program

Hierarchies of geometries for shape were proposed by Felix Klein in a famous speech at the University of Erlangen in 1872. Klein argued that different geometries can be stratified by the *properties of objects* that are *preserved* by different types of transforms (Klein 1893).

The category of transforms, Klein suggested to consider, was the *group of motions*, i.e., 1-to-1 mappings of a space onto itself, that leave intact certain properties of objects in the considered space.

Norman and Todd have argued that (Norman & Todd 1992: 106):

There is a growing amount of evidence to suggest that human vision involves a similar type of stratification in which the most perceptually salient aspects of an object’s structure are those that remain invariant over the largest number of possible transforms.

For an illustration of this concept of a hierarchy of geometries for a planar surface refer to Lord and Wilson’s treatise (Lord & Wilson 1984: Fig. 7, p. 35).

Remark 3.7 (Differential geometry) *When one studies geometry in a local neighbourhood m of a smooth manifold M , one can reduce such local probing of differentiable structure to infinitesimal interrogations in the tangent space to $m \subset M$, i.e., the first order, linear, approximation to M at m ; this permits to “reduce” geometry to linear algebra.*

3.2.1 Riemannian geometries

A Riemannian geometry is a generalisation of the intrinsic geometry of a surface.²¹ The associated (metric) spaces are spaces of *constant curvature*. If the (Riemannian or Gaussian) curvature of the space is zero, then we have a *Euclidean* space which is

homogeneous and isotropic. For strictly positive curvature, we have a *spherical* or *elliptic* space (e.g., the space for “flat inhabitants” of the Earth). Finally, for strictly negative curvature we have the *hyperbolic* space of J. Bolyai and N. I. Lobachevski.

Hyperbolic geometry. Study of those properties of subsets of $\mathbb{R}^n$ which are invariant under the application of any hyperbolic transform; i.e., a transform based on a hyperbolic metric, such as the Minkowski distance, defined within a hyperbolic space (i.e., a space with negative intrinsic curvature).

It has been proposed in the literature that the correct geometry of binocular vision or stereopsis is the hyperbolic one (Luneburg 1947, 1950). The latter gives an explanation for some well-known perceptual illusions, such as, e.g., a perceptually straight line being physically (*i.e.*, in Euclidean – environmental – space) concave toward the observer (Blank 1978).

Euclidean geometry. Study of those properties of subsets of $\mathbb{R}^n$ which are invariant under the application of any Euclidean transform; i.e., a “distance-preserving” transform based on a Euclidean metric within a Euclidean space (i.e., a space with vanishing intrinsic curvature). Thus, concepts such as angles and lengths are meaningful.

Consider the example of a “rigid motion” in the plane $\mathbb{R}^2$, an instance of a Euclidean transform. Then, the outline of the transformed object is preserved. In this sense it is a “very rigid” geometry, since any slight deformation of a form tells us that we are dealing with a different object.

In particular, a circle is different from an ellipse in Euclidean geometry. So is a square with respect to a parallelogram.

Norman and Todd emphasise that (Norman & Todd 1992:95):

Throughout the literature on human perception, classical Euclidean geometry is by far the most common framework for describing the structure of the environment.

From Euclidean geometry, we may “descend” in the hierarchy to reach *affine geometry* which can be portrayed as Euclidean geometry without a metric or without a “machinery to measure distances and angles” (Koenderink 1990:58).

3.2.2 Non-metrical geometries

Affine Geometry. Study of those properties of subsets of $\mathbb{R}^n$ which are invariant under the application of any 1-to-1 affine transform (e.g., a transform which preserves parallelism and ratios of length along any line).

Consider the example of a “shear motion” in the plane $\mathbb{R}^2$. The outline is not strictly preserved anymore. A square is transformed into a parallelogram. A circle into an ellipse. As such, squares and circles are not generic objects in affine geometry.

try: they do not define a class of objects being invariant under an affine transform. In fact the concepts of squares and circles make no sense in affine geometry; but parallelograms and ellipses do.

Paraphrasing Norman and Todd (Norman & Todd 1992:96):

In affine geometry the distance metric is allowed to vary in different directions – i.e., it is anisotropic. Arbitrary stretching transformations, which do not leave distance invariant, but do preserve other properties such as the sign of Gaussian curvature of a surface or the parallelism of pair of line segments, are permitted in this geometry.

In stereopsis, the mapping between the two views has been modelled as an affine transform, to recover the gradient of horizontal disparity (Jones & Malik 1992).

Projective geometry. Study of those properties of subsets of $\mathbb{R}^n$ which are invariant under the application of any projective transform such as a perspective.

E.g., under a general projective transform, a conic section remains a conic section, and a collinear set of points remains collinear.

In comparison to affine geometry where all *parallelograms* are equivalent, in 2D projective geometry, all *quadrilaterals* are equivalent, i.e., parallelism is not necessarily preserved.

Ordinal geometry. Study of those properties of subsets of $\mathbb{R}^n$ which are invariant under the application of a non-metric transform such that only simple order relations, such as “greater than” or “lesser than,” are preserved.

James Gibson has argued that some of our perceptual awareness of the environment is based on such a geometry (Gibson 1950). Todd and Reichel have suggested that an observer’s knowledge of smoothly curved surfaces can often involve a form of ordinal representation such that neighbouring surface patches are labelled in terms of which region is qualitatively “closer” to the point of observation (Todd & Reichel 1989). Note that, occlusion contours are an important source of information in such a geometry.

A general deformation of a continuum where ordinal structure is preserved can provide a vast class of transforms for such a geometry (see the paragraph on “General deformation,” §3.1.6, and the paragraph below on “rubber” geometry).

3.2.3 *Pseudo-metrical geometries*

These are geometries which can accommodate a metric but do not absolutely require one.

“Rubber” geometry (Kendall’s morphometrics). Study of those properties of subsets of $\mathbb{R}^n$ which are invariant under the application of any general deformation, i.e., a *continuous* and *differentiable* 1-to-1 mapping.

A famous application of this type of geometry into the field of *morphometrics* was made in the early 20th century by D’Arcy Wentworth Thompson in his study of the evolutionary changes of shapes of certain animals and plants (Thompson 1992).²²

In stereopsis, the mapping between the two views has been modelled as a deformation, to recover the horizontal disparity (Koenderink & van Doorn 1976).

If one gets rid of the requirement of differentiability and only imposes continuity, one gets a “topological geometry.”

Topological geometry. Study of those properties of subsets of $\mathbb{R}^n$ which are invariant under the application of any “topological transform,” such as a *homotopy*. The latter may be loosely described as a family of continuous maps which permits to “continuously deform” a class of sets covering two objects so that they coincide.

For example, imagine having two sets of curves covering two surfaces (our two objects’ outline here): a homotopy will establish a (continuous) correspondence between these two covers.²³ Therefore, two objects are considered topologically equivalent, and thus having the same shape, if a deformation exist which can map one to the other and vice versa. E.g., a sphere and a cube are topologically the same object, but not a sphere/cube and a torus.

Note that it has been observed that “topologically equivalent” shapes “produce stronger perceptions of apparent motion” (Norman & Todd 1992: 105).

A special class of topological geometry concerns sets which are self-similar, such as the famous “monster curves” and “fractals.”

Fractal geometry. Study of those properties of subsets of $\mathbb{R}^n$ which are invariant under the application of any topological transform and which present the characteristic of *self-similarity*. Three main forms of self-similarity are considered: (i) *iterated transforms*, which have a fixed geometric replacement rule, e.g., the Koch snowflake which produces an infinite perimeter enclosing a finite area; (ii) *recurrence transforms*, which define an iterative sequence at each point in space, e.g., the Mandelbrot set (Mandelbrot 1982); (iii) *random transforms*, which use a stochastic sequence rather than a deterministic one, e.g., the fractal landscapes of Ken Musgrave (Ebert et al. 2002) or the Lévy flights, a type of random walks (Mandelbrot 1982).

4. Horizontal and vertical transform sequences

Having looked at different transforms for shape in isolation, including the “classical” geometry stratification based on invariance as proposed by Klein, I now propose a more general bi-partite classification for *sequences of transforms* which are applied to the outline and various shape representations of an object. This classification which *does away with the invariance constraint*, is inspired in particular by the early filters developed in Mathematical Morphology, based on idempotence, by the works of Zucker et al. on the application of differential transforms for the recovery of shape, and by Michael Leyton’s work on the association of a history for shape, based on those properties which *vary* under deformations.

A transform sequence involves at least a pair of transforms, where, typically, the first one acts directly on the outline of an object, resulting in a first representation, while the second one then acts on this representation to make more explicit special selected types of features.

I will distinguish such concatenations of transforms as to whether they operate (i) along the outline of a shape or tangentially, or (ii) away from it or perpendicularly. I call the two classes “horizontal” and “vertical” transform sequences. This represents a fundamental dichotomy in the analysis of shape transforms. While the horizontal class is inspired by a more classical approach to geometry based on calculus and the study of those properties which remain intrinsic to the object, i.e., such that the analysis can be carried on (or along) the outline itself, independently of the surrounding (or embedding) space, the vertical class is inspired by more modern thinking in mathematics, such as singularity theory, and is based on an analysis of the outline under transformation together with its surrounding space.

In the following I provide a number of examples from the field of computer vision which illustrate how the two classes are used to solve similar problems.

4.1 Horizontal transform sequences

Definition 4.1 (Horizontal transform sequence) *It is a shape transform sequence, such that the first transform operates in the tangent space (or the tangent bundle) of the outline of the object; in other words, subsequent transforms are constrained to act along the outline of the object.*

Example 4.1 (1-D manifold on the plane: Co-circularity mapping and curvature) *Let a tangent field on a curve segment, $\alpha : I \subset \mathbb{R} \rightarrow \mathbb{R}^2$ be defined; where this curve segment represents, e.g., part of the outline (boundary) of a planar object (or of its projection in the image plane).*

Tangents are co-circular if they lie on a common circular arc of radius r . So one can take a transform from the tangent space to the co-circularity space for pairs of tangents.²⁴

Furthermore, co-circularity is directly related to the concept of curvature as one takes tangents being closer and closer to each other. In the limit, one can define a real map on the interval, I , used to parameterised the curve to define the concept of curvature, k , a scalar field:

$$k : I \rightarrow \mathbf{R}, \quad k = \lim_{\Delta s \rightarrow 0} \frac{\Delta \theta}{\Delta s} = \frac{1}{r},$$

where $s \in I$ is the parameter and θ is the angle between the tangents (or their corresponding normals). Thus, curvature measures deviation from flatness, or collinearity, for successive tangents.

The corresponding (limiting) circle is called the osculating circle (Pogorelov 1958:57). It has second order contact with the curve α , i.e., both the curve and the circle agree (locally) in position, tangency and curvature (or up to second derivatives). If s is the arc-length parameter, then k is equal to the norm (length) of the acceleration vector α'' (Pogorelov 1958:50):

$$k(s) = \|\alpha''(s)\|.$$

In conclusion, if a circle exists such that a pair of tangents (to α) are also tangent to this circle, then they are said to be co-circular. The (osculating) circle is thus used to “connect” or relate the pair of tangents in a “horizontal manner,” i.e., along the curve through a simple interpolating shape, a circle, or through a derived “co-circularity support,” i.e., a “directed” neighbourhood or thick trace (Parent & Zucker 1989).

Note that, co-circularity is related to Blum’s notion of *symmetric chord coordinates* (Blum 1973) and Brady and Asada’s *local symmetries* (Brady & Asada 1984). Instead of relating the pair of tangents through an arc of circle, link them by a straight line segment. If they are co-circular then this line segment is a *chord* (w.r. to the arc of circle) and the angles each tangent makes with respect to this chord are equal (e.g., see Parent & Zucker 1989: Fig. 5(b)). The extension to surfaces involves the classical notion of “parallel transport” from differential geometry.

Example 4.2 (2-D manifold: Parallel transport) Let V be a tangent vector field over a region B of the surface M . How can we relate (or move) a given tangent vector $\mathbf{v}$ at point $\mathbf{p} \in M$ with another tangent vector $\mathbf{w}$ at point $\mathbf{q} \in M$? One possibility is through the notion of (Levi-Civita) parallel transport where a tangent vector, e.g., $\mathbf{v}$, is moved along a curve $\alpha \subset M$ such that it stays parallel with respect to the tangent space T_M . Thus, $\mathbf{v}$ may rotate, but only with T_M (this depending on the intrinsic or Gaussian curvature of M).

In computer vision, such a notion of parallel transport has been used by Sander and Zucker in the context of the recovery of the trace of surface points (Sander & Zucker 1990).²⁵ But, note that a surface can curve in more than one way at any given point. In particular, for smooth surfaces one can show that at any point of the manifold (normal) curvatures range from a minimum, $k_{\min}$, to a maximum, $k_{\max}$, called principal curvatures. Thus, the object used for relating tangent vectors on M cannot simply be an osculating sphere (the extension to 3D of the 2D osculating circle above), but rather needs to be a more flexible object. In the context of differential geometry, the “canonical” object used is the osculating paraboloid, S , a parabolic quadric surface having second order contact with the manifold M (Pogorelov 1958: 87).

Therefore, we can compare/relate tangent vectors on M through parallel transport via an interpolating paraboloid. We need the latter because we do not know *a-priori* the exact “form” (trace) of M . One usually has only partial/noisy knowledge of the trace of M . This is an instance of a “horizontal transform sequence.”

Remark 4.1 (Inadequacy of parallel transport) *Sander notes that parallel transport is not the “ideal” way of relating tangent vectors of a surface for the following reasons (Sander 1988: 38–39):*

- *Which curve/path to choose on the paraboloid? There is potentially an infinity of them and, although parallel transport is well-defined for a given curve (i.e., it is unique), it gives, in general (when $K \neq 0$), different results for different curves. I.e., the holonomy angle varies with the path, and this is a function of the Gaussian (intrinsic) curvature of the paraboloid.*
- *How to transport “directions” (associated with a given initial tangent vector, $\mathbf{v}$, and its corresponding normal curvature) so as to preserve them? Clearly, arbitrary paths will not preserve such directions. An interesting choice might be lines of (principal) curvature, but, then there is the problem of “navigating” through singularities of the tangent field, i.e., through umbilics (points where all curvatures are equal: $k_{\min} = k_{\max}$).*

Sander proposes the notion of “compatibilities via frame bundles” to alleviate these problems (Sander 1988).²⁶

Example 4.3 (2-D manifold: Compatibilities via frame bundles) *The local information available in a cross section, σ , of the frame bundle via the first and second fundamental forms, I_S and II_S ,²⁷ of the interpolating paraboloid, S , enables the direct generation of the frame $\xi_\alpha(\mathbf{q})$ corresponding to moving ξ_α from $\mathbf{p}$ to $\mathbf{q}$.²⁸ This moving frame will tell “what the local information at” $\mathbf{q}$ “looks like according to the information at” $\mathbf{p}$ (Sander 1988: 41). Thus, it gives “support” to the information at $\mathbf{q}$, “under the assumption of an underlying smooth surface” (i.e., S).*

This is again an *horizontal* transform sequence where frames are “moved” according to an interpolating surface which best approximates (through some optimisation measure) the actual surface M we wish to recover (the trace of).

Example 4.4 (2-D manifold: “Classical” Shape from Shading (SfS)) *Shape from shading consists in “inferring geometric structure of a scene from measurements of its photometric structure” (Langer & Zucker 1994).*

The “classical” model studied in the context of this problem, is based on the “Image Irradiance Equation” (IIE) which states that luminance, or image intensity I , is proportional to some reflectance function or map, $\mathcal{R} : I(\mathbf{x}) = \mathcal{R}(\mathbf{N}(\mathbf{x}))$, where $\mathcal{R}$ is assumed dependent on the surface normal, $\mathbf{N}$, i.e., dependent on the surface orientation. Here, the surface trace is the geometric structure one wishes to recover from the image data, i.e., it is an “inverse optics problem.” A special case (the “classical” one) is when the surface is assumed Lambertian,²⁹ i.e., the light source is assumed punctual and remote (far), and the albedo (the amount of light reflected, a material property), ρ , is assumed constant over the surface. Then, the IIE can be expressed as (Langer & Zucker 1994):

$$I(\mathbf{x}) = \rho I_s \mathbf{L} \cdot \mathbf{N}(\mathbf{x}) + \rho I_a,$$

where $\mathbf{L}$ and I_s are the direction and intensity of the distant point source, respectively, and I_a is the intensity of the ambient illumination. The (constant) term ρI_a is usually ignored in the calculations. The term $\rho I_s \mathbf{L}$ is assumed known (or is approximated empirically). Then, assuming one faces a Monge patch, i.e., the surface topography can be represented as the graph of a function, $z = f(x, y)$, which is always possible locally for smooth surfaces, the (unit) normal vector $\mathbf{N}$ can be expressed as a function of the gradient of the surface, i.e., in terms of the slopes, $p = \partial z / \partial x$ and $q = \partial z / \partial y$, of the surface in the x and y directions, respectively:

$$\mathbf{N} \cdot \|\mathbf{N}\| = \mathbf{v}^x \wedge \mathbf{v}^y = (1, 0, p) \wedge (0, 1, q) = (-p, -q, 1),$$

where the vectors $\mathbf{v}^x$ and $\mathbf{v}^y$ are tangent vectors to the surface in the x and y directions. Similarly, for a distant source, we can represent its direction $\mathbf{L}$ via a unit vector $(-p^s, -q^s, 1)$ expressed in terms of the slopes p^s, q^s , of a (wave) surface perpendicular to the light ray. Thus the IIE, in the Lambertian–single–distant–light–source case, reduces to (Horn 1986: 219):

$$I(x, y) = \mathcal{R}(p, q) = \frac{1 + p^s \cdot p + q^s \cdot q}{\sqrt{1 + p^2 + q^2}},$$

which has been normalised by a factor ρI_0 . This equation can be seen as a nonlinear first order PDE.

Different ways of solving numerically the reduced IIE equation above have been used in the computer vision literature. All of these are instances of *horizontal* transform sequences where one, starting from an initial contour, usually taken as an

extrema of I , grows solutions *along the surface* by (numerically) integrating the IIE. Essentially, one takes small steps from a known height/position³⁰ of the surface and recovers surface height at the new position.³¹ To “validate” the calculations one can integrate solutions on the calculated surface, this along closed paths from and to the initial (known) starting position; thus, the difference in height should vanish.³²

Example 4.5 (Motion field & optical flow) *One may consider a motion (vector) field as the “object” of interest. The layout of the field is transformed, e.g., under assumptions of smoothness, thus retrieving a smooth vector field possibly corresponding to some translation and/or rotation of a real physical object in the scene.*³³

*A common hypothesis made in computer vision is that the motion field, which assigns a velocity vector to each point in an image, corresponds to the optical flow which is the apparent motion of the brightness pattern.*³⁴ *Note that this correspondence is an idealisation; in many (practical) situations, the motion field and the optical flow are not equivalent. Nevertheless, optical flow is believed to be a potentially important source of information towards the retrieval of the true motion field or of some of its features.*

*Under the (very restrictive) assumption that brightness varies smoothly one can derive the Optical Flow Constraint Equation (OFCE) (Horn 1986:282), which is nothing but the 1-jet for I :*³⁵

$$I_x u + I_y v + I_t = 0 ,$$

where $u = dx/dt$ and $v = dy/dt$ are the optical flow components, and the (partial) derivatives I_x , I_y and I_t are estimated from the image. We may re-write this equation as follows:

$$\frac{\partial I}{\partial t} + \nabla I \cdot (u, v) = 0 .$$

This can be shown to correspond to a conservation law for an incompressible medium, where the image intensity, I , corresponds to the density of the medium (Fitzpatrick 1988). In this context, the assumption of incompressibility for a density image is analogous to the assumption that image brightness of a surface point is independent of the motion of that point (a very restrictive assumption indeed).

Note that the OFCE is a *local* constraint on brightness velocities. In order to recover a “consistent” (global) optical flow, one has to filter or transform somehow the initial flow field given by the values (u, v) at each image point. E.g., one may convolve the field with an averaging kernel. Or one may update information on the basis of the available local geometric structure (i.e., tangent direction, curvature) within a relaxation network. The latter is reminiscent of ex. 4.3, where compatibilities are (here) in terms of *tangent bundles*, i.e., the problem can be cast as one of selecting consistent sections σ through the tangent bundle of the “motion surface”

(i.e., motion field). Thus the flow field is integrated along the underlying surface (or “base”), and we can picture it as an horizontal transform sequence.³⁶

I covered in the above a number of illustrative examples of horizontal transform sequences from the field of computer vision, from contour tracing to shape from shading and shape from motion. In the next section, I consider the “complementary” vertical ones, where the transforms operate in the space away from the outline of an object, rather than along it.

4.2 Vertical transform sequences

Definition 4.2 (Vertical transform sequence) *It is a shape transform sequence, such that the first transform operates in the normal space (or the normal bundle) of the outline of the object; in other words, subsequent transforms are constrained to act above (or below – but not along, i.e., acting away from) the outline of the object.*

Example 4.6 (Medial symmetry graphs (MSG) in 2D) *There exists a few different but equivalent or related definitions of the concept of a medial symmetry set of a 2D shape, taking the form of a skeletal graph.³⁷ One interest in having a variety of definitions is that each one leads to different operational schemes for implementing the transform sequence. Below we give the main different definitions to be found in the literature.*

Contact circles: *Consider the closure of the loci of centres of disks tangent to the outline at two or more points; such bitangent disks are called “contact circles.” The centres of these circles trace a symmetry set ($\mathcal{SS}$) which in 2D takes the form of a planar graph. The most commonly used subset of the $\mathcal{SS}$ is the medial axis ($\mathcal{MA}$) of Harry Blum (Blum 1973) for which all contact circles are constrained to be maximal, i.e., such that they do not contain any points of the outline other than the contact (tangency) points.³⁸*

Quenching waves: *Imagine a wave propagating from the boundary with a constant speed. Shocks occur where wavefronts merge, intersect or collapse. The locus of each shock correspond to the centre of a contact circle part of the $\mathcal{SS}$, since such a point is at equal distance from the two contact points on the outline, generators of the pair of quenching waves. Or imagine a fire propagating over a grassfield and initiated by the outline of the object. The loci of quenching points of the grassfire correspond to Blum’s $\mathcal{MA}$. That is, if we let waves intersect and keep propagating we get the $\mathcal{SS}$, and if we terminate propagation as soon as waves create a first shock we get the $\mathcal{MA}$.*

Voronoi diagrams and cut locus: *Consider the set of points being closest to a given outline sample. This is the natural framework to work with the Voronoi Diagram ($\mathcal{VD}$) (Okabe et al. 2000), a complete partition of space made of cells whose boundary, the Voronoi Graph, is equivalent to the $\mathcal{MA}$. Each cell (or Voronoi*

region, $\mathcal{VR}$) is made from the loci of the $\mathbb{R}^2$ plane closest to one single outline point. A related concept is that of the Cut Locus, the closure of the set containing all points that have at least two shortest segments to the outline (Wolter 1992).

Ridges of a distance surface: Consider the surface generated by applying a distance transform to the object's form. Map distance as a height function. Then consider the loci of points on ridges of this topographical map (Kotelly 1963; Blum 1973; Leymarie & Levine 1992b).

Mid-chord symmetry and process inferring symmetry axis (PISA): An alternative to the centre of contact circles, as is used to trace the $\mathcal{MA}$, is to consider instead (i) the mid-chord point defined between the two contact points on the outline (Blum 1973; Brady & Asada 1984), or (ii) the mid-point of the shortest subtended arc of the circle (Leyton 1987).

Shock graph: Consider the radius function associated to the contact circles. It defines a (1D) vector field along the trace of the $\mathcal{MA}$. The singularities of this function when linked by the intermediary $\mathcal{MA}$ segments define a more compact graph representation of the $\mathcal{MA}$. These singularities, the sources, relays and sinks of the radius function, correspond to endpoints, junctions and saddle points of the $\mathcal{MA}$ (Kimia 2003).

Now let us consider the boundary of the object as either a piecewise 1D manifold or a discrete set of samples. Then the MSG is a transform which maps this boundary onto the medial symmetry set in one of the forms given above. This transform acts along normal rays to the boundary of the shape. That is, it acts along the *normal bundle*. Points where normal rays meet correspond to symmetry points. This is an instance of a vertical transform sequence.

Concept 4.1 (Wolter's topological shape theorem of the $\mathcal{MA}$) *Franz-Erich Wolter has shown, in his early works, how the $\mathcal{MA}$ of a solid O (i.e., such that the outline defines an inside and an outside for the object under scrutiny) "contains the essence of the topological type" of O (Wolter 1992; Wolter et al. 2004). In summary, Wolter has proven for the cases of smooth outlines and, in 2D, for piecewise-smooth outlines, that the $\mathcal{MA}$ is the deformation retract of the solid O ; and, therefore, it preserves the homotopy type of O , which can be retrieved, for example, from the (topological) graph structure of the $\mathcal{MA}$.*

Remark 4.2 (Relation between the MSG and co-circularity mapping) *In the case of the MSG, two points are put in relation with each other by searching where their normal rays meet. This locus defines the contact circle.*

In the case of the co-circularity mapping, the osculating circle is used to relate pairs of tangents of boundary points by using the circular shape as an interpolant (ex. 4.1, §4.1).

Thus, the same object, a circle in close contact with the outline, is used in two different manners, for relating pair of points of the boundary which share a “local symmetry,” albeit of different nature.

Remark 4.3 (Planar and space curves) *In the above we considered only outlines tracing planar curves in $\mathbb{R}^2$. The discussion also applies to space curves in $\mathbb{R}^3$, although the normal bundle is then more complex: it involves both the normals and bi-normals (Bruce & Giblin 1992).*

Example 4.7 (MSG on curved surfaces) *Wolter et al. have extended the concept of a medial symmetry graph from the $\mathbb{R}^2$ plane to curved surfaces embedded in $\mathbb{R}^3$ (Rausch et al. 1996; Kunze et al. 1997; Wolter & Frieese 2000). They use the notion of the cut locus for this purpose, where shortest paths are now defined along shortest geodesics from an outline defined on the curved surface.³⁹ In essence, this is an extension from the Euclidean space (flat 2D manifolds, i.e., with zero curvature) to the Riemannian space (curved 2D manifolds).*

The above notion of medial symmetry sets for 1D outlines is naturally extended to 2D outlines, i.e., to the traces of surfaces of 3D objects.

Example 4.8 (Medial symmetry sets in 3D) *Medial symmetries of 3D objects can be captured by replacing contact disks by contact spheres.*

Again, one may view these medial symmetries in different flavours. For example, a wave might propagate from the surface boundary along the normal rays to the boundary. Singularities of this normal mapping correspond to the symmetry set. Most of the definitions used for the 2D case have been extended to 3D, with the exceptions of mid-chord symmetries and Leyton’s PISA.

Remark 4.4 (Medial symmetries in 3D and compatibilities via frame bundles) *In both cases we are putting in relation pairs of surface points. In the case of medial symmetries we do so by navigating along normal rays to the surface, while in Sander and Zucker’s horizontal transform sequence one navigates along the surface according to an interpolating paraboloid.*

Note that different objects are used to relate pairs of points: osculating paraboloids in the horizontal transform sequence of Sander and Zucker and contact spheres for the “vertical” medial symmetries.

Note also that, the “compatibilities via frame bundles” are used in practice to relate surface points within a relatively small neighbourhood along the surface – a kind of “sausage” neighbourhood also called “thick-trace” – while in the vertical case, much larger (3D) neighbourhoods may be used to compute symmetries. In principle, all the available region of space under scrutiny is used.

The main difference between medial symmetry sets in 2D and 3D, is that, in 3D, for outlines sampling the surfaces of volumetric object, the resulting set is not a graph in general, but rather a hypergraph made of intersecting medial surfaces; in other words the medial set has one more dimension. To benefit from the use of a graph representation, e.g., in recognition tasks, a number of additional transforms are available.

Example 4.9 (Medial symmetry graphs in 3D) *Medial symmetries of 3D objects constrained to take the form of a 3D (non-necessarily planar) graph. Below we give the main definitions to be found in the literature.*

Generalised cylinders and 3D curve skeletons: *A generalised cylinder is defined by (i) an axis, (ii) a radius functional, often called “sweep function,” and (iii) a set of rules to define branching points and behaviours. The axis is typically a 3D space curve. The radius functional specifies the trace of the outline perpendicularly to the axis, and may vary as it “sweeps” along the axis.⁴⁰ The set of rules defines what happens at the end points of an axis: in which directions other axes with their own radius functional may branch-out. While generalised cylinders are useful in generating pre-specified objects, 3D curve skeletons are an attempt to define from the outline an equivalent representation. These come under different names, representing slightly different definitions which we organise in two main categories: (i) curve skeletons (Borgefors et al. 1999), skeleton graphs (Sundar et al. 2003) or ik-skeletons (Wade & Parent 2002) are obtained by thinning a discrete distance field, such as resulting from an object voxelisation, and (ii) Reeb graphs (Hilaga et al. 2001) also called Level Set Diagrams (Verroust & Lazarus 2000) are obtained by a simulated flooding of the object as a solid, tracing midpoints of level sets as the object is filled-in.*

Medial and shock scaffolds: *The 3D MA of Blum has been generalised by Leymarie, Kimia and Giblin by defining a unique 3D graph structure from which the MA itself can be recovered (Leymarie & Kimia 2001; Leymarie 2003; Giblin & Kimia 2004; Leymarie & Kimia 2007). This is an extension of the concept of a shock graph in 2D and it comes into two main types: (i) the Medial Scaffold (MS) which links the singularities of the radius function associated to the MA for two types of medial curves: axial shock curves where three or more MA surfaces intersect, and rib curves which are associated with surface ridges of the object’s outline; (ii) the Shock Scaffold (SC) which extends the MS by defining additional curves linking the singularities of the radius function along the MA surfaces.*

Critical nets and molecular graphs: *Consider each 3D Voronoi region, $\mathcal{VR}$, of the Voronoi diagram, $\mathcal{VD}$, and its decomposition in a set of k -faces which are homeomorphic to a closed ball, i.e., such that each $\mathcal{VR}$ can be continuously deformed to a closed ball of same dimensions (Bredon 1993). The collection gives a finite*

closed covering of $\mathbb{R}^n$ implying that the $\mathcal{VD}$ is a CW-complex representation of space.⁴¹ Another classical result, from Morse theory, a branch of differential topology (Milnor 1963), associates to each of the Voronoi k -face a critical point of index $n - k$ for a given Morse function, which we can in practice take as a function of the distance to the outline.⁴² The graph obtained by connecting in this “ $n+1$ -level hierarchy” the critical points, seen as nodes, via links which are unique topological “separatrices” is called a critical net in Crystallography (Johnson et al. 1996; Johnson 1999), and a molecular graph in Computational Chemistry (Bader 1990: Ch. 2 & 3); such graphs are used to describe atomic structures, e.g., crystals or proteins, and the topology of their charge density.⁴³

Generalised cylinders and 3D curve skeletons are well-adapted to capture the shape of tubular-like objects. When describing more general objects in 3D, the scaffolds, $\mathcal{MS}$ or $\mathcal{SC}$, are to be preferred. While, in theory, the Critical nets and Molecular graphs can be seen as an extension of the $\mathcal{SC}$, they have only been studied for the restricted case where the outline is sampling an outward surface bounding a density cloud centred on atomic centres of molecules, and for non-degenerate underlying Morse functions, constraints not imposed on the $\mathcal{MA}$ and its representation as a graph via the $\mathcal{MS}$ or the $\mathcal{SC}$.

Similarly to the 2D MSG case, all of the above transform sequences in 3D operate away from the outline, along the normal rays to the object’s underlying surface: they are all instances of vertical transform sequences.

Example 4.10 (Shape from shading: Case of diffuse light source conditions)

Paraphrasing from the work of Langer and Zucker (Langer & Zucker 1994):

Under diffuse lighting conditions, surface luminance depends primarily on the amount of sky visible from each surface patch, while surface normal and mutual illumination are of secondary importance. This dependence – on the visible sky – is modelled in terms of a surface aperture function: the percentage of incident directions visible in a local hemisphere above “ground” or above a surface patch. An approximately linear constraint between image intensity and surface aperture is found.

While in the case of “classical SfS” (§4.1), where one assumes a point light source (at infinity), the local geometry constraints lie along the surface, in the case of the diffuse lighting problem, the local geometric constraints are found in the ambient light above the surface.

In this “SfS on a cloudy day,” once surface apertures are estimated from an initial image, a recursive marching algorithm is used to refine depths for a viewer assumed seeing the scene from above. The entire space is discretized via an Eulerian (fixed) grid and depth is propagated one node at a time; the method resembles a blanket slowly falling on the scene which starts folding locally according to the

surface aperture values. All computations take place in the ambient space above the object's surface, yet another instance of a vertical transform sequence.

Example 4.11 (Biological motion field: Compact medial node representation)

Ilona Kovács et al. have proposed that special medial nodes along a thick trace of the MA represent a plausible natural and effective mechanism for the computation of motion fields in animal visual systems. A pseudo-distance function, D_ϵ , is defined which captures medial symmetries within a pair of disks defining an annulus region, where ϵ denotes the thickness of the annulus. Under this metric, the special nodes of the modified MA are those which locally maximise the amount of outline trace they capture: they are “the most informative points along the skeleton” (MA) of an object in motion (Kovács et al. 1998). Such a compact medial representation is inspired by the chronophotographs of Etienne-Jules Marey (circa 1880) and G. Johansson's work (circa 1970) on the perception of biological motion for point-light walker displays.

The D_ϵ distance function of Kovács et al. is computed for each frame of a video sequence, and can be applied in each frame directly to the available outlines. It is comparable to other “annular” medial symmetry schemes (Kelly & Levine 1995; van Tonder & Ejima 2000, 2003), and this group also falls in the class of vertical transform sequences. Note that as ϵ goes to zero, the method may converge to the shock graph representation, where the special medial nodes along the MA are those for which the associated distance (or radius) function goes through a critical point (i.e., a source, relay or sink).

4.3 Mixed transform sequences

It is possible to define sequences of transforms which *mix horizontality and verticality*. Typically, this proves useful when one considers instantiation of general deformations. An important example of such mixed transform sequences, which we briefly consider below, is the so called reaction-diffusion case.

4.3.1 Reaction-diffusion transform sequence

Definition 4.3 (Reaction-diffusion transform sequence (RDTs)) *Transform sequence which operates in both the normal and tangent space of the object's outline on the basis of a differential equation. The action of the RDTs is equivalent to a combined normal flow – the “reaction” part – and a tangential flow – the “diffusion” part.*

Example 4.12 (Curve evolution) *Let the evolution of a piecewise continuous 2D outline be modelled by a differential equation of the form:*

$$\frac{\partial \alpha}{\partial t} = \Upsilon(\cdot) \mathbf{N},$$

where $\alpha(s, t) = (x(s, t), y(s, t))$ is the position vector of the outline, s is the length parameter (not necessarily arclength), $Y(\cdot)$ is the deformation functional, and $\mathbf{N}$ is the unit normal vector (typically oriented “inward” with respect to the object’s outline).

The following deformation functional: $Y \equiv \beta_0 + \beta_1 \kappa$, where κ stands for the curvature of a curve segment α , is often considered in differential geometry (curve shortening and smoothing), physics of interfaces, and computer vision (Catté et al. 1992; Kimia et al. 1995; Sethian 1999).

When $\beta_0 = 0$ the evolution reduces to a pure diffusion (horizontal) transform sequence: a curvature driven evolution which shrinks the curve (to a circular dot for a closed outline). When $\beta_1 = 0$ the evolution reduces to a pure reaction (vertical) transform sequence: an eikonal wave propagation for the outline developing shocks (*MA* symmetries). When both effects are active (i.e., $\beta_0 \neq 0$ and $\beta_1 \neq 0$) shape is visualised as a scale-space of deformations (Kimia et al. 1995). How to explore this scale space is neither systematic, nor obvious. Furthermore, it depends on two “parameters” (β_0 and β_1) which may be adapted to different applications.

Example 4.13 (Alan Turing’s morphogens) *The most famous RDTS was proposed by Alan Turing (circa 1950) to model the changing spatial concentrations of chemical substances he called “morphogens” (Turing 1952). By varying the relative influence of reaction versus diffusion, different regular patterns observable in nature have been simulated, such as the stripes of zebras and the spots of leopards (Murray 2004).*

Having introduced in the above a classification of transform sequences based on the dichotomy of their action along the outline or away from it, I now consider in the next section systemic definitions of transform sequences in the form of “grammars:” i.e., sets of rules permitting to group, assemble, separate, re-organise, object elements with the goal of capturing complexity in forms.

5. Grammars for shape

Definition 5.1 (Shape grammar) (i) *It consists in a set of rules which permits to replace one outline by another, by giving a “shape–interpolation,” i.e., a plausible developmental history.* (ii) *Alternatively, it is a set of rules which permits to add/delete (or cut and paste) parts/components to an outline as a function of its derived shape.*

An example of the former is Leyton’s process-grammar (§5.1) and an example of the latter is Richards and Hoffman’s codon-grammar (§5.2), both of which will be described below.

Thus objects of various complexities may be generated by an iteration of a finite number of simple outline transformation instructions. As such, a shape

grammar provides a *discrete form* of a language which can be based on shape transforms.

Various shape grammars have been developed in the literature. E.g., the generation of self-similar fractal objects is possible with very simple grammars (Lord & Wilson 1984: §8.1). Selection rules, which forbid the addition of a subunit under certain conditions, have been used by Ulam to generate less regular patterns (Lord & Wilson 1984: §8.2). Trees and river systems, crystals, tessellations and space filling organisations are other examples of domain of applications of shape grammars as object generators (Lord & Wilson 1984). An important example of early work is to be found in “L-systems,” also called “Lindenmayer systems” or “parallel string-rewrite systems,” which are made from productions rules used to define a tracing of piecewise linear segments with joints parameterised by rotation angles (Lindenmayer 1968; Prusinkiewicz & Lindenmayer 1999). These rules also are a compact way to iteratively repeat constructive sequences in the description of fractals, often used to model groups of plants, flowers, leaves, and so on (Ferraro et al. 2005).

Remark 5.1 (The inverse problem of vision) *Note that, for the representation and description of an object, toward its recognition, a shape grammar will prove particularly useful if it is possible to “invert” the rules, or at least identify them, when given a particular outline of an object. This would then permit to identify the process which gave the “final” shape, or alternatively would give cues in identifying parts or object components.*

Example 5.1 (Biederman’s recognition-by-components theory) *The RBC theory of object recognition by Biederman et al. is an important example of a shape grammar which aims at addressing the “inverse problem” of vision (Biederman 1987; Hummel & Biederman 1992; Biederman 1995). The RBC theory starts from the hypothesis that a finite compact family of (36) object parts (or components) called geons is sufficient to represent most commonly encountered objects. Geons include primitive parts (or geometric units) such as blocks, cylinders, spheres, which are parameterised by a number of (“non-accidental”) features encoding size, tapering, symmetry axes, bends, and so on. Geons can be extracted or re-constructed from outline traces in images. Their sequencing in representing a complex object is specified by relations (geometrical and topological) amongst parts.*

Since its proposal in the late 1980’s the RBC theory has been implemented in a few systems aiming at performing object recognition. Bergevin and Levine implemented the first comprehensive system based on RBC called PARVO (Primal Access Recognition of Visual Objects) (Bergevin & Levine 1993). Wu and Levine restricted Biederman’s set of 36 geons to only 7 defined as parameterised subsets of superellipsoids for potentially faster processing (Wu & Levine 1997). Of related scope is the earlier work by Pentland et al. using parameterised object primitives to

model parts (Pentland & Scarloff 1991). An account of the various approaches and issues in using geons until the mid-1990's can be found in (Dickinson et al. 1997). More recent work can be found in (Siddiqi & Kimia 1995; Siddiqi et al. 2001; Sing & Hoffman 2001).

Remark 5.2 (Dynamic shape and genetic grammars) *One can generalise shape grammars within the context of cellular automata where some randomisation is introduced in the manifestation of the “rules” leading to “dynamic” shapes; for example see the works of Wolfram et al. (Wolfram 1994) and more recent studies in biological pattern genesis (Deutsch & Dormann 2005).*

Another possible generalisation is in the context of genetic programming where mutations and the natural mixing of a pool of “genes” (possibly representing shape components or features) is used to obtain evolving “natural” shapes; for example see the early works of Dawkins on biomorphs (Dawkins 1986), and of Latham and Todd on genetic art (Todd & Latham 1992), and more recent works in arts and design (Bentley 1999) and biological sciences (Kumar & Bentley 2003).

In the following I consider and compare two important grammars which emerged in the field of visual perception analysis in the 1980's, this to illustrate some of the issues one has to face in designing a “shape language.”

5.1 Leyton's process–grammar

Leyton's starts from the assumption that “shape [can be] understood as the outcome of [physical] processes that formed it” (Leyton 1988). Thus, a shape is “defined as the outcome of a history which is psychologically decomposed into *phases* where qualitatively different processes were acting;” and, “categories of objects are psychologically stratified into levels corresponding to phases in a formation history.”

Leyton has looked at two types of process–history: when considering the object as a whole, and when considering it only through a few salient features – i.e., the curvature extrema of the outline. In the following I only consider the latter approach.

For smooth, planar, simply connected outlines, Leyton shows that a simple grammar of *six* operations is sufficient to represent the relationships between any two shapes.

In order to retrieve the processes which are (potentially) acting on a shape, Leyton suggests that it is necessary to recover the *curvature extrema* of the bounding contour of an object and its associated *symmetry axes*. The inference of processes from a single outline requires two stages

1. From *curvature extrema* to *symmetry axes*.
2. From *symmetry axes* to *processes*.

Curvature extrema and symmetry axes are put into correspondence through the following principle.⁴⁴

Theorem 5.1 (2D symmetry–curvature duality (Leyton 1987)) *Any segment of a smooth planar curve bounded by two consecutive curvature extrema of the same type (i.e., max. or min.) has a unique differential symmetry axis which terminates at a curvature extremum of the opposite type (i.e., min. or max., respectively).*⁴⁵

Thus, this “duality” between curvature extrema and symmetry axes gives a rule for relating them. Furthermore, symmetry axes are postulated as the “*records of deformational processes*.” E.g., the longer an axis is, the more significant the attached process shall be. A process in this context may be thought of as an internal or external force acting on the shape. Significance may be taken in terms of magnitude or duration in time.

Proposition 5.1 (Interaction principle (Leyton 1984)) *The symmetry axes of a perceptual organisation are interpreted as the principal directions along which processes are most likely to act or have acted.*

This proposition was justified as follows by Leyton (Leyton 1985, 1988):

A transformation, acting on an organisation, is one in which symmetry axes become invariant lines or eigenspaces under the transformation. Thus the transformation will tend to preserve the symmetries – i.e., be *structure-preserving* on the organisation.

By combining the above two principles one gets:

Corollary 5.1 ((Leyton 1988)) *Each curvature extrema implies a process whose trace is the unique symmetry axis associated with, and terminating at, that extremum.*

Thus, a process explains the extrema.

Corollary 5.2 (Asymmetry rule (Leyton 1988)) *Processes are understood as creating greater curvature variation, i.e., greater information in the sense of mathematical information theory.*⁴⁶

Remark 5.3 (Simplest shape is the circle) *If one uses the “asymmetry rule” above to extrapolate backward in time, the “ultimate starting point for the shape must have been the circle.” The latter possess the least amount of curvature variation, hence information.*

Definition 5.2 (Types of curvature extrema) *Four types of curvature extrema are defined: local minimum and maximum, denoted by m and M , respectively, each being*

possibly of negative (–) or positive (+) curvature. The semantic interpretation of these extrema is as follows:

M^+ :	“Protrusion”	m^- :	“Indentation”
m^+ :	“Squashing”	M^- :	“Internal resistance”

Leyton’s Process–grammar

The process–grammar consists of two types of *extrapolation rules*: Continuation and Bifurcation rules, denoted by C and B , respectively. Continuation implies that no new extremum is created during the (deformation) process (but an extremum may change in nature), while a bifurcation implies that a new, intervening, extremum must be introduced. The (rewrite) rules of the grammar can be summarised as follows:

$CM^+ :$	$M^+ \rightarrow 0M^+0$	“Protrusion continues”
$Cm^- :$	$m^- \rightarrow 0m^-0$	“Indentation continues”
$Cm^+ :$	$m^+ \rightarrow 0m^-0$	“Squashing continues until it indents”
$CM^- :$	$M^- \rightarrow 0M^+0$	“Internal resistance continues until it protrudes”
$BM^+ :$	$M^+ \rightarrow M^+m^+M^+$	“Protrusion bifurcates”
$Bm^- :$	$m^- \rightarrow m^-M^-m^-$	“Indentation bifurcates”
$Bm^+ :$	$m^+ \rightarrow m^+M^+m^+$	“Protrusion introduced”
$BM^- :$	$M^- \rightarrow M^-m^-M^-$	“Indentation introduced”

NB: Continuation at protrusion, M^+ , and indentation, m^- , are structurally trivial: i.e., no change in the extremum’s nature occurs. As such they do not need be considered explicitly in the grammar. Thus, only the last six rules are to be used in practice.

By applying the rules of the above process–grammar to simple objects, such as an ellipse, one can generate a succession of more complex outlines. Note that, all outlines can in principle be derived from the simplest one in the hierarchy. It turns out that the *shape–space* so generated is highly structured (e.g., see Figures 16 and 17 in Leyton (1988)). However, there exists more than one “path,” in general, through this shape–space, when travelling from a simple outline to a more complex one. Leyton suggested the following simple heuristic for constraining the search in this space:

Heuristic 5.1 (Size–is–time (Leyton 1988, 1989)) *In the absence of information to the contrary, size corresponds to time. I.e., later processes have a shorter time to develop than earlier ones and thus in general they tend to result in smaller deformations of an outline.*

Leyton also suggested that *deblurring* of the boundary (coordinates) could provide a way of finding a single interpretation or path through the shape-space (Leyton 1989). Leyton claims that deblurring is the appropriate shape transform to use because it “embodies the size-is-time heuristic;” and in particular: “deblurring incrementally moves the boundary in the direction of increased curvature variation.”

Remark 5.4 (Figure-ground reversal) *Figure-ground reversal is simply dealt with by introducing a duality operation, D , which interchanges extrema labels as follows:*

$$\begin{array}{llll} M^+ & \leftrightarrow & m^- & : \text{ protrusion} \leftrightarrow \text{ indentation} \\ M^- & \leftrightarrow & m^+ & : \text{ squashing} \leftrightarrow \text{ internal resistance} \end{array}$$

5.2 Richards and Hoffman’s codon-grammar

This is an extrema-grammar rather than a process-grammar, i.e., it is based only on the curvature extrema of the object’s outline (Richards & Hoffman 1985).

Definition 5.3 (Codon) *A boundary segment made of a triple of extrema, such that two are minima, m , and are endpoints for the segment.*

There are five types of non-trivial codons for planar, smooth, and simply-connected curve segments: $\{m^+M^+m^+, m^-M^-m^-, m^+M^+m^-, m^-M^-m^+, m^-M^+m^-\}$. These are denoted $0^+, 0^-, 1^+, 1^-, 2$, respectively (Richards & Hoffman 1985) – i.e., where the numeral represents the number of zeroes of curvature present in the curve segment.

If one considers their duals, to account for figure-ground reversal, one gets a family of ten codons (Leyton 1988).

Codon-grammar

The codon-grammar consist of two types of rules:

Level-preserving: Replace a codon by a codon.

Level-increasing: Replace an extremum with a codon or replace a codon with a pair of consecutive codons.

The codon-grammar consists of eighteen rewrite rules (see Leyton 1988: Tbl. 2 for a list of these).

Remark 5.5 (Comparison of the process and codon grammars)

1. *The Process-grammar is more economical: only six rules are necessary in comparison of the eighteen rules of the Codon-grammar.*
2. *Every codon rule can be represented by a sequence of Process-grammar rules.*

3. *The codon substitutions are cutting and pasting operations, while the Process-grammar rewrite rules are developmental, i.e., concerned with growth.*

Note that both the Process and Codon grammars can be implemented on the basis of horizontal or vertical transform sequences, to identify curvature extrema and their relationships. An example of an horizontal approach is the curvature morphology system of Leymarie and Levine which maps an outline to a set of valleys and peaks of curvature where the scale of an extremum is characterised by the size of a valley (mapping to a concavity for the outline) or a peak (a convexity for the outline) (Leymarie & Levine 1989). An example of a vertical approach is the work of Tek and Kimia based on the 2D *MA* where the collection of axes can be used to rank order a set of curvature extrema (sitting at one end of each axis) by a notion of significance or scale: how important each extremum is in relation to other nearby ones and the overall shape (Tek & Kimia 2001). Significance is directly associated to symmetry axes in how long they are or how large an object's sub-part they capture; as such this system provides an implementation of some of Leyton's ideas, and in particular his "size is time" heuristic (5.1).

An important open question when designing a practical shape language, is whether a *discrete* form of a family of transforms for shape is more adequate than a *continuous* one, such as those provided by the Reaction-Diffusion Transform Sequences (RDTS; §4.3.1). Note that a discrete, rule-based approach provided by a shape grammar such as Leyton's or the Codon system, has the appeal of being more easily implementable in a reasoning system; e.g., via an expert-system. Doing reasoning in a continuous shape-space derived from an approach such as the RDTS does not appear, at first glance, as easy, since the space is not explicitly as well "*compartmentalised*." A similar issue occurs in neuroscience where the leading paradigm is one where *perception* relies on discrete time events and processing (VanRullen & Koch 2003).

6. Conclusion

In the above "thoughts" I have tried to characterise the meaning of what "shape" is and how it can be related to fundamental mathematical concepts, via the notion of transforms: mathematical constructs permitting to characterise the probing of the structural field of objects, thus revealing their shape. I have also given many examples based on my understanding of visual perception and computer vision, to illustrate and better focus the effort. A particular emphasis was the definition of a bi-partite classification of sequences of transforms, whether they first act *along* the outline or *away* from it: what I call *horizontal* and *vertical* transform sequences. This dichotomy is useful as a mean to compare and distinguish popular shape

transforms, as was illustrated with examples from the field of computer vision, and it could be extended to other “visual” fields. For example, in CAD-CAM (Patrikalakis & Maekawa 2002), one can distinguish methods based on the differential geometry of curves and surfaces as well as of intersections – essentially sets of horizontal transform sequences – from the methods based on offsets for curves and surfaces – i.e., vertical transform sequences. How the horizontal versus vertical approaches precisely compare, when applied to similar or identical problems – e.g., when identifying corners or curvature extrema – is a study which remains to be conducted. Furthermore, such a dichotomy may also be useful to study the human visual system itself. The hypercolumn system of receptive fields first identified by Hubel and Wiesel (Heeger et al. 1996) can be understood as the basis of a (differential) geometric engine used to generate tangent bundles, curvature maps, and so on, from outline samples (Zucker 1987); in other words, we can portrait (at least part of) the hypercolumn system of the visual cortex as implementing certain horizontal transform sequences. Kovács et al. have observed in human subjects differential contrast sensitivity maps for 2D shapes which are consistent with a medial function representing the percentage of outline samples equidistant from the observation point within a tolerance level (Kovács et al. 1998). This has lead them to hypothesize a medial *MA*-based shape processing method for human vision, an instance of vertical transform sequences.

This “essay” is also the beginning of a survey in the context of visual analysis with the goal of addressing the very nature of shape understanding. Most topics surveyed in this manuscript would benefit from a more detailed description, with illustrative examples. Also, some topics have only been touched upon: in particular the last topic of shape grammars, and the genesis of shape, which would benefit from a comparative study of the different approaches now available in computing, such as “shock grammars” (Siddiqi & Kimia 1996; Sebastian et al. 2004). The latter can be seen as recent developments in computing inspired from the original approach to morphogenesis developed by the mathematician René Thom in the 1970’s, the founder of catastrophe theory (Thom 1989), which lead to the emergence of singularity theory and its applications (Arnold 1989; Bruce & Giblin 1992; Poston & Stewart 1996).

Notes

1. One can also talk of the shape of a region of space, even if empty ($\varphi = 0$), but defined by some outline: e.g., the region of space in between objects such as the voids between galaxies in the universe (Weygaert 2003; Jones et al. 2005).
2. Leyton talks about “transfer structure” rather than “structural field,” notions I consider equivalent for the purpose of my somewhat imprecise study.

3. Refer to Leyton for an explanation of how *memory* and *shape* can be tied up together (Leyton 2001).
4. The Hough transform maps edgels of a binary image to a parameter space. In the case of a square one would use the slope and intercept of the equation of a straight line as the parameters. Edgels which fall on nearby line segments will cluster in the parameter space.
5. CAD-CAM: Computer Aided Design and Manufacturing.
6. This is similar to the *Hough Transform* where lines in a binary image are mapped to points in the *parameter space*.
7. A “bias” in favour of invertible transforms for shape may have emerged in the signal processing field, whose original concerns about (lossless) data transmission are very relevant, e.g., for communication systems.
8. The structural element is typically defined as a set of pixels or voxels, similarly to the input dataset (an image array). Set operations, in particular Minkowski’s addition and subtraction, are then used to construct higher order operators.
9. E.g., this rules out the well-known Marr–Hildreth detector of zero-crossings of the Laplacian of an image I , which creates “shadow edges,” i.e., edges of edges, if applied twice.
10. Unless we change the constraint set or perturb the data set, of course – which is what one may try to do to reach a global minimum.
11. *Adapted boundaries*: hyper-surfaces on which one of the coordinate is *constant*.
12. This is a restriction of a conformal map (i.e., which preserves local angles) which in the literature is applied as a local constraint to more general functions, such as in complex analysis.
13. A real square matrix A is positive definite if $\mathbf{v}^T A \mathbf{v} > 0$. This is a useful property to solve a linear system of equations, $A\mathbf{x} = \mathbf{y}$ by a (Cholesky) decomposition in a (upper) triangular matrix.
14. For more details refer to Arnold’s (Arnold 1989:Ch. 8) and Guillemin’s (Guillemin & Sternberg 1984) manuscripts.
15. For a good introduction on this topic, refer to Gurtin’s manuscript (Gurtin 1982).
16. The original statement gives a pair of solutions in general, only one of which is of practical interest: the envelope at an increasing distance from the original sources of propagation.
17. In differential geometry, $\mathbf{v}$ is called the *gauge figure* or *indicatrix*: i.e., the set of (possible) velocity vectors in every direction at $\mathbf{q} \in E^3$ (Koenderink 1990). In general, $\mathbf{v}$ can vary in space, which proves necessary to model propagation at interfaces or in non-homogeneous media. In the domain of visual perception, a varying index could be used to model contextual propagation to favour certain orientations versus others (like horizontal versus vertical proximity relationships between boundary elements).
18. See also the work of Kotelly on Blum’s transform (Kotelly 1963). For a more recent application of the distance surface $\tau = \Phi(x, y)$ to the 2D skeletonization problem, when simulating a grassfire on a square lattice, see the work of Leymarie and Levine (Leymarie & Levine 1992b).
19. Another way of thinking of this concept, is in term of the system of normals to smooth surfaces in Euclidean space. In neighbourhoods of such smooth surfaces, their normals (rays) form a smooth fibration – they constitute tightly packed fibres (beams) filling-in space, hence creating a regular distance hypersurface – but, at some distance from the surface, the various normals begin to intersect one another (Arnold 1989:480).

20. A real function $f : I \rightarrow \mathbb{R}$ such that $\exists$ a constant $K \geq 0$, such that $|f(x) - f(y)| \leq K |x - y|$ for all x, y in the interval I , is called a Lipschitz continuous function: i.e., it is limited in *how fast it can change*.
21. Historical note: This generalisation of the concept of Euclidean geometry was proposed by Bernhard Riemann in 1854 in his lecture to qualify for the post of lecturer in Gauss's department at the University of Göttingen: "Über die Hypothesen welche der Geometrie zu Grunde liegen" (Lord & Wilson 1984).
22. For a more recent account on the field of morphometrics, refer to the works of Fred Bookstein et al. (1992, 1997).
23. Two (mathematical) objects are said to be homotopic if one can be continuously deformed into the other (after Henri Poincaré, circa 1900 (Collins 2004)).
24. For the "classical" application to the computer vision problem of the recovery of the trace of curve points in an image, refer to the work of Parent & Zucker (1989).
25. This work can be seen as an extension to surfaces of the ideas of Parent and Zucker developed for curves (Parent & Zucker 1989).
26. A "frame bundle" on a manifold M is a set of orthonormal (coordinate) frames defined at every point $p \in M$: it determines the possible choices for an orthonormal basis of the tangent space of M , T_M ; it also reflects the ambiguity of choosing coordinates on M . The choice of a (local) cross section of the frame bundle gives a "moving frame" used to calculate "tensors" such as curvatures.
27. The first fundamental form defines the inner product on the tangent space (of pairs of tangent vectors): $I_S(\mathbf{v}, \mathbf{u}) = \mathbf{v} \cdot \mathbf{u}$. The second fundamental form defines a symmetric bilinear form on the tangent space: $II_S(\mathbf{v}, \mathbf{u}) = \mathfrak{Z}(\mathbf{v}) \cdot \mathbf{u}$, where $\mathfrak{Z}$ is the "shape operator" giving the (negative) derivative of the unit normal vector field of a surface (a kind of curvature). The ratio II_S/I_S computes normal curvature to the surface.
28. A frame on a 2-D manifold M is an orthonormal coordinate system, such that two of its axes span the tangent space T_M at any given point $\mathbf{p} \in M$. Note that at each point of M , $\exists$ an infinite number of such frames (an equivalence class) through rotation around the normal third axis. Hence the need for the notion of *bundle* of frames.
29. "An ideal Lambertian surface is one that appears equally bright from all viewing directions and reflects all incident light, absorbing none" (Horn 1986:212). The luminance of a Lambertian surface depends only on the angle between the surface normal and the incident light ray.
30. For $I(x, y) = I_{\max}$ one can set the height $z = z_{\max}$ (i.e., either the brightest or darkest spot in a image) to some arbitrary constant.
31. Refer to Berthold Horn's manuscript for more details on the methods of "characteristic strips" and "variational relaxation" for solving this classical shape from shading problem (Horn 1986: Ch. 11).
32. This is an example of the Jacobi identity for Lie transforms (§3.1.5).
33. For work along these lines, where a geometrical image transformation is put in correspondence with the physical motion of the imaged objects, see e.g., the work of Fitzpatrick (1988).
34. Historical note: Gibson et al. coined the term "optical flow;" e.g., refer to (Gibson et al. 1959) for some of their early ideas on the subject.

35. A n -jet is a Taylor series expansion of which one keeps only the terms up-to (and including) n^{th} -order derivatives.
36. This is an extension to the optical flow problem of the ideas on “shading flows and scenel bundles” of Zucker et al. (Breton et al. 1992).
37. I.e., it resembles a stick figure or skeleton for an anthropomorphic form.
38. Using the “closure” requirement in the definition insures to capture limit points as ends of $\mathcal{SS}$ branches. For example, in computing the $\mathcal{MA}$ of an ellipse, this ensures the two focal points are included as end points.
39. The result can as well be “stated by using maximal geodesic circles, [making it a closer cousin] to the Euclidean maximal circle concept” (private communication from F.-E. Wolter).
40. The origins of generalised cylinders can be traced back at least to the 18th century geometer Gaspard Monge who defines “pipe” surfaces as the envelope of the set of spheres of radius r centred on a space curve defining the axis or “spine” of the pipe (Monge 1850). Pipe surfaces are used in CAD-CAM for shape reconstruction, blending of surfaces, surface transitions, numerically-controlled (NC) verification (Patrikalakis & Maekawa 2002: 353).
41. A CW-complex, i.e., a “closure finite weak topology cell complex,” is a finite collection of non-empty regions called “cells” homeomorphic to a closed ball, such that (i) the cells have pairwise disjoint interiors, and (ii) the boundary of each cell is the union of other cells (of lower dimensions) (Edelsbrunner & Shah 1997). CW-complices, a key tool in algebraic topology (Bredon 1993), are a “nice” class of topological spaces which are general in scope and simple in construction.
42. *Critical points* of a function f on a manifold M are found where the gradient vanishes, i.e., $\nabla f = 0$. Such a function is said to be “Morse” if its critical point are non-degenerate, i.e., they are isolated.
43. Such “molecular” shape descriptions have recently started being used in Computer Graphics, for modelling objects via *meta-balls*, that is, networks of spheres of varying radii, connected into a 3D graph representation (Hart 1999).
44. This was first proved only in the case of *regular, simply connected, planar curve*. It was then extended to deal with the more interesting case of first order discontinuities, such as corners, creases and cusps, in Hayes & Leyton (1989).
45. The 3D version of this theorem states that “At each curvature extremum on a *principle line of curvature*, there is a unique *symmetry sheet* that terminates at that extremum” (Yuille & Leyton 1990). Technically speaking, this argument is always valid (for 2D and 3D problems) when considering the symmetry set ($\mathcal{SS}$) (Bruce & Giblin 1992; Giblin & Kimia 2004). When considering only the $\mathcal{MA}$ (subset of the $\mathcal{SS}$) some axes may be missing in correspondence to surface ridges, i.e., lines of curvature extrema of the boundary. Recall that the $\mathcal{MA}$ sheets terminate at centres of curvature associated to ridges, rather than on the bounding surface itself.
46. Here information is identified with *variety*. Refer to Howard Resnikoff who has shown that the *measurement of curvature variation has the same form as the measurement of information* in information theory (Resnikoff 1989).

References

- Arnheim, R. (1974). *Art and Visual Perception: A Psychology of the Creative Eye*. Berkeley, CA: University of California Press. New version; expanded and revised edition of the 1954 original.
- Arnold, V. I. (1989). *Mathematical Methods of Classical Mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer-Verlag, 2nd edition.
- Bader, R. F. W. (1990). *Atoms in Molecules – A Quantum Theory*, volume 22 of *International Series of Monographs on Chemistry*. Oxford, UK: Clarendon Press.
- Bamberg, P., & Sternberg, S. (1988). *A Course in Mathematics for Students of Physics*, volume 1. Cambridge University Press.
- Bentley, P. J. (Ed.). (1999). *Evolutionary Design by Computers*. Morgan Kaufmann.
- Bergevin, R., & Levine, M. D. (1993). Generic object recognition: Building and matching coarse descriptions from line drawings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15, (1), 19–36.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115–147.
- Biederman, I. (1995). Visual object recognition. In S. F. Kosslyn & D. N. Osherson (Eds.), *An Invitation to Cognitive Science*, volume 2 of *Visual Cognition*, Chapter 4 (pp. 121–165). MIT Press.
- Blank, A. A. (1978). Metric geometry in human binocular perception: Theory and fact. In E. L. J. Leeuwenberg & H. F. J. M. Buffart (Eds.), *Formal Theories of Visual Perception*, Chapter 4 (pp. 83–102). John Wiley and Sons.
- Blum, H. (1967). A transformation for extracting new descriptors of shape. In W. Wathen-Dunn (Ed.), *Models for the Perception of Speech and Visual Form* (pp. 362–380). Cambridge, MA: MIT Press. Proceedings of a Symposium held in Boston, MA, November 1964.
- Blum, H. (1973). Biological shape and visual science. *Journal of Theoretical Biology*, 38, 205–287.
- Bookstein, F. L. (1992). *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press.
- Bookstein, F. L. (1997). Landmark methods for forms without landmarks: Morphometrics of group differences in outline shape. *Medical Image Analysis*, 1(3), 225–243.
- Borgefors, G. et al. (1999). Computing skeletons in 3D. *Pattern Recognition*, 32(7), 1225–1236.
- Brady, M., & Asada, H. (1984). Smoothed local symmetries and their implementation. *International Journal of Robotic Research*, 3(3), 36–61.
- Bredon, G. E. (1993). *Topology and Geometry*, Volume 139 of *Graduate Texts in Mathematics*. Springer-Verlag.
- Breton, P., Iverson, L., Langer, M., & Zucker, S. W. (1992). Shading flows and scene bundles: A new approach to shape from shading. In G. Sandini (Ed.), *Proceedings of the 2nd European Conference on Computer Vision (ECCV)*, Volume 588 of *Lecture Notes in Computer Science* (pp. 135–150). Santa Margherita, Ligure, Italy. Springer-Verlag.
- Bruce, J. W., & Giblin, P. J. (1992). *Curves and Singularities: A Geometrical Introduction to Singularity Theory*, 2nd edition. Cambridge University Press, UK.
- Burt, P. J. (1981). Fast filter transforms for image processing. *Computer Graphics and Image Processing (CGIP)*, 16, 20–51.
- Burt, P. J. (1983). Fast algorithms for estimating local image properties. *Computer Vision, Graphics and Image Processing (CVGIP)*, 21, 368–382.

- Catté, F., Lions, P.-L., Morel, J.-M., & Coll, T. (1992). Image selective smoothing and edge detection by nonlinear diffusion. *SIAM Journal on Numerical Analysis*, 29(1), 182–193.
- Collins, G. P. (2004). The shapes of space. *Scientific American*, 291, 94–103.
- Danielsson, P.-E. (1980). Euclidean distance mapping. *Computer Graphics and Image Processing (CGIP)*, 14, 227–248.
- Dawkins, R. (1986). *The Blind Watchmaker*. W. W. Norton and Company. Reissued in 1990 (UK) and 1996 (USA).
- Deutsch, A., & Dormann, S. (2005). *Cellular Automaton Modeling of Biological Pattern Formation*. Modeling and Simulation in Science, Engineering and Technology. Birkhäuser.
- Dickinson, S. J., Bergevin, R., Biederman, L., Eklundh, J.-O., Munck-Fairwood, R., Jain, A. K., & Pentland, A. (1997). Panel report: The potential of geons for generic 3-D object recognition. *Image and Vision Computing*, 15(4), 277–292.
- Ebert, D. S., Musgrave, F. K., Peachey, D., Perlin, K., & Worley, S. (2002). *Texturing and Modeling – A Procedural Approach*. Morgan Kaufmann, 3rd edition.
- Edelsbrunner, H., & Shah, N. R. (1997). Triangulating topological spaces. *International Journal of Computational Geometry and Applications*, 7(4), 365–378.
- Ferraro, M., & Caelli, T. M. (1988). Relationships between integral transform invariances and Lie group theory. *Journal of the Optical Society of America – A*, 5(5), 738–742.
- Ferraro, P., Godin, C., & Prusinkiewicz, P. (2005). Toward a quantification of self-similarity in plants. *Fractals*, 13(2), 91–109.
- Fitzpatrick, J. M. (1988). The existence of geometrical density image transformations corresponding to object motion. *Computer Vision, Graphics and Image Processing (CVGIP)*, 44(2), 155–174.
- Giblin, P. J., & Kimia, B. B. (2004). A formal classification of 3D medial axis points and their local geometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 26(2), 238–251.
- Gibson, E. J., Gibson, J. J., Smith, O. W., & Flock, H. (1959). Motion parallax as a determinant of perceived depth. *Journal of Experimental Psychology*, 8(1), 40–51.
- Gibson, J. J. (1950). *The Perception of the Visual World*. Boston, MA: Houghton Mifflin Cie. Reprinted in 1974 by Greenwood Press, Westport, CT.
- Goodman, A. W. (1964). A partial differential equation and parallel plane curves. *American Mathematical Monthly*, 71, 257–264.
- Guillemin, V., & Sternberg, S. (1984). *Symplectic Techniques in Physics*. Cambridge University Press, UK.
- Gurtin, M. E. (1982). *An Introduction to Continuum Mechanics*, volume 158 of *Mathematics in Science and Engineering Series*. Harcourt Publishers.
- Hansen, O. (1992). *On the Use of Local Symmetries in Image Analysis and Computer Vision*. PhD thesis, Laboratory of Image Analysis, Institute of Electronic Systems, Aalborg University, Denmark.
- Hart, J. C. (1999). Using the CW-complex to represent the topological structure of implicit surfaces and solids. In *Proc. Implicit Surfaces '99* (pp. 107–112). Eurographics/SIGGRAPH, ACM.
- Hayes, P. J., & Leyton, M. (1989). Processes at discontinuities. In *Proceedings of the 11th Joint Conference on Artificial Intelligence*, volume 2 (pp. 1267–1272). Detroit, MI: American Association for A.I.
- Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (1996). Computational models of cortical visual processing. In *Proceedings of the National Academy of Science*, volume 93 (pp. 623–627). Washington, DC.

- Hilaga, M., Shinagawa, Y., Kohmura, T., & Kunii, T. L. (2001). Topology matching for fully automatic similarity estimation of 3D shapes. In *Proc. SIGGRAPH 2001* (pp. 203–212). Los Angeles, CA: ACM.
- Hoffman, W. C. (1966). The Lie algebra of visual perception. *Journal of Mathematical Psychology*, 3, 65–98. Errata in vol. 4, pp. 348–349, 1967.
- Hoffman, W. C. (1978). The Lie transformation group approach to visual neuropsychology. In E. L. J. Leeuwenberg & H. F. J. Buffart M. (Eds.), *Formal Theories of Visual Perception*, Chapter 2 (pp. 27–66). John Wiley and Sons.
- Hoffman, W. C. (1984). Figural synthesis by vectorfields: Geometric neuropsychology. In P. C. Dodwell & T. Caelli (Eds.), *Figural Synthesis*, Chapter 8 (pp. 249–282). Lawrence Erlbaum Associates.
- Horn, B. K. P. (1986). *Robot Vision*. Electrical Engineering and Computer Science Series. Cambridge, MA: MIT Press.
- Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, 99, 480–517.
- Johnson, C. K. (1999). Crystallographic topology 2: Overview and work in progress. In V. Alexiades & G. Siopsis (Eds.), *Trends in Mathematical Physics*, Studies in Advanced Mathematics. AMS, Providence, RI.
- Johnson, C. K., Burnett, M. N., & Dunbar, W. D. (1996). Crystallographic topology and its applications. Oak Ridge National Laboratory. Preprint available: <http://www.ornl.gov/sci/ortep/topology/preprint.html>
- Jones, B. J. T., Martínez, V. J., Saar, E., & Trimble, V. (2005). Scaling laws in the distribution of galaxies. *Reviews of Modern Physics*, 76, 1211–1267.
- Jones, D. G., & Malik, J. (1992). Determining three dimensional shape from orientation and spatial frequency disparities. In G. Sandini (Ed.), *Proceedings of the 2nd European Conference on Computer Vision (ECCV)*, volume 588 of *Lecture Notes in Computer Science* (pp. 661–669). Santa Margherita, Ligure, Italy: Springer-Verlag.
- Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision (IJCV)*, 1(4), 321–331.
- Kelly, M. F., & Levine, M. D. (1995). Annular symmetry operators: A method for locating and describing objects. In *Proceedings of the 5th International Conference on Computer Vision (ICCV)* (pp. 1016–1021).
- Kimia, B. B. (1991). *Toward a Computational Theory of Shape*. PhD thesis, Electrical Engineering, McGill University, Montreal, Canada.
- Kimia, B. B. (2003). On the role of medial geometry in human vision. *Journal of Physiology*, 97(2–3), 155–190.
- Kimia, B. B., Tannenbaum, A. R., & Zucker, S. W. (1995). Shapes, shocks, and deformations I: The components of two-dimensional shape and the reaction-diffusion space. *International Journal of Computer Vision*, 15(3), 189–224.
- Klein, F. (1893). Vergleichende Betrachtungen über neuere Geometrische Forschungen (“Erlangen Programm”). *Mathematische Annalen*, 43, 63–100.
- Koenderink, J. J. (1990). *Solid Shape*. A.I. Series. Cambridge, MA: MIT Press.
- Koenderink, J. J., & van Doorn, A. J. (1976). Geometry of binocular vision and a model for stereopsis. *Biological Cybernetics*, 21(1), 29–35.
- Kotelly, J. C. (1963). A mathematical model of Blum’s theory of pattern recognition. Research report AFCRL-63-164, U.S. Air Force Cambridge Research Laboratories.
- Kovács, L., Feher, A., & Julesz, B. (1998). Medial-point description of shape: A representation for action coding and its psychological correlates. *Vision Research*, 38, 2323–2333.

- Kumar, S., & Bentley, P. J. (Eds.). (2003). *On Growth, Form and Computers*. Academic Press.
- Kunze, R., Wolter, F.-E., & Rausch, T. (1997). Geodesic Voronoi diagrams on parametric surfaces. In *CGI '97* (pp. 230–237). IEEE, Computer Society Press.
- Lagarde, J. W. (1990). Constraints and their satisfaction in the recovery of local surface structure. Master's thesis, Electrical Engineering, McGill University, Montreal, Canada.
- Langer, M. S., & Zucker, S. W. (1994). Shape-from-shading on a cloudy day. *Journal of the Optical Society of America – A*, 11(2), 467–478.
- Leymarie, F. (1990). Tracking and describing deformable objects using active contour models. Technical report CIM-90-9, McGill University, Dept. of Electrical Engineering, Montreal, Canada. www.lems.brown.edu/~leymarie/meng/.
- Leymarie, F., & Levine, M. D. (1989). Shape features using curvature morphology. In W. A. Pearlman (Ed.), *Proc. of the SPIE Conf. on Visual Communications and Image Processing IV*, volume 1199 (pp. 390–401).
- Leymarie, F., & Levine, M. D. (1992a). Fast raster scan distance propagation on the discrete rectangular lattice. *Computer Vision, Graphics and Image Processing – Image Understanding (CVGIP-IU)*, 55(1), 84–94.
- Leymarie, F., & Levine, M. D. (1992b). Simulating the grassfire transform using an active contour model. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 14(1), 56–75.
- Leymarie, F., & Levine, M. D. (1993). Tracking deformable objects in the plane using an active contour model. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 15(6), 617–634.
- Leymarie, F. F. (2003). *Three-Dimensional Shape Representation via Shock Flows*. PhD thesis, Brown University, Providence, RI, USA. www.lems.brown.edu/~leymarie/phd/.
- Leymarie, F. F., & Kimia, B. B. (2000). Discrete 3D wave propagation for computing morphological operations from surface patches and unorganized points. In J. Goutsias, L. Vincent, & D. Bloomberg (Eds.), *Math. Morphology and its Applications to Image and Signal Processing*, volume 18 of *Computational Imaging and Vision Series* (pp. 351–360). Kluwer Academic.
- Leymarie, F. F., & Kimia, B. B. (2001). The shock scaffold for representing 3D shape. In C. Arcelli, L. P. Cordella, G. S. d. B. (Eds.), *Visual Form*, number LNCS 2059 (pp. 216–229). Springer-Verlag.
- Leymarie, F. F., & Kimia, B. B. (2007). The medial scaffold of 3D unorganized point clouds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Accepted for publication.
- Leyton, M. (1984). Perceptual organization as nested control. *Biological Cybernetics*, 51, 141–153.
- Leyton, M. (1985). Generative systems of analyzers. *Computer Vision, Graphics and Image Processing (CVGIP)*, 31(2), 201–241.
- Leyton, M. (1987). Symmetry-curvature duality. *Computer Vision, Graphics and Image Processing (CVGIP)*, 38, 327–341.
- Leyton, M. (1988). A process grammar for shape. *Artificial Intelligence Journal*, 34(2), 213–247.
- Leyton, M. (1989). Inferring causal history from shape. *Cognitive Science*, 13, 357–387.
- Leyton, M. (2001). *A Generative Theory of Shape*. Lecture Notes Computer Science Number 2145. Springer-Verlag.
- Lindenmayer, A. (1968). Mathematical models for cellular interactions in development: Parts I and II. *Journal of Theoretical Biology*, 18, 280–315.
- Lord, E. A., & Wilson, C. B. (1984). *The Mathematical Description of Shape and Form*. Ellis Horwood Series in Mathematics and its Applications. Statistics and Operational Research. Halsted Press.

- Luneburg, R. K. (1947). *Mathematical Analysis of Binocular Vision*. Princeton, NJ: Princeton University Press. Published for the Hanover (Dartmouth Eye) Institute, N.H., U.S.A.
- Luneburg, R. K. (1950). The metric of binocular visual space. *Journal of the Optical Society of America (JOSA)*, 40(10), 627–642.
- Mandelbrot, B. (1982). *The Fractal Geometry of Nature*. W. H. Freeman & Co.
- Milnor, J. W. (1963). Morse theory. In *Annals of Mathematics Studies*, volume 51. Princeton, NJ: Princeton University Press.
- Monge, G. (1850). *Application de l'Analyse à la Géométrie*. Bachelier, Paris, France.
- Montanari, U. (1969). Continuous skeletons from digitized images. *Journal of the Association for Computing Machinery*, 16(4), 534–549.
- Murray, J. D. (2004). *Mathematical Biology*, 3rd edition, in 2 volumes. Springer.
- Norman, J. F., & Todd, J. T. (1992). The visual perception of 3-dimensional form. In G. A. Carpenter & S. Grossberg (Eds.), *Neural Networks for Vision and Image Processing*, Chapter 4 (pp. 93–110). Cambridge, MA: MIT Press.
- Okabe, A., Boots, B., Sugihara, K., & Chiu, S. N. (2000). *Spatial tessellations: Concepts and applications of Voronoi diagrams*. Probability and Statistics series, 2nd edition. Wiley.
- Owens, R., Venkatesh, S., & Ross, J. (1989). Edge detection is a projection. *Pattern Recognition Letters*, 9(4), 233–244.
- Papathomas, T. V., & Julesz, B. (1989). Lie differential operators in animal and machine vision. In J. C. Simon (Ed.), *From Pixels to Features* (pp. 115–126). North-Holland. Proceedings of a Workshop held at Bonas, France, 22–27 August 1988.
- Parent, P., & Zucker, S. W. (1989). Trace inference, curvature consistency and curve detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 11(8), 823–839.
- Patrikialakis, N. M., & Maekawa, T. (2002). *Shape Interrogation for Computer Aided Design and Manufacturing*. Springer.
- Pentland, A., & Scarloff, S. (1991). Closed-form solutions for physically based shape modeling and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 13(7), 715–729.
- Perlick, V. (2004). Gravitational lensing from a spacetime perspective. *Living Reviews in Relativity*, 7(9). <http://www.livingreviews.org/lrr-2004-9>
- Pogorelov, A. V. (1958). *Differential Geometry*. P. Noordhoff, Groningen, The Netherlands.
- Poston, T., & Stewart, I. (1996). *Catastrophe Theory and Its Applications*, new edition. Dover.
- Prusinkiewicz, P., & Lindenmayer, A. (1999). *The Algorithmic Beauty of Plants*, 2nd edition. Springer-Verlag.
- Rausch, T., Wolter, F.-E., & Sniehotta, O. (1996). Computation of medial curves in surfaces. In *Conference on the Mathematics of Surfaces VII* (pp. 43–68). Institute of Mathematics and its Applications (IMA).
- Resnikoff, H. L. (1989). *The Illusion of Reality: Topics in Information Science*. Springer-Verlag.
- Richards, W., & Hoffman, D. D. (1985). Codon constraints on closed 2D shapes. *Computer Vision, Graphics and Image Processing (CVGIP)*, 31, 265–281.
- Ronse, C. (1993). On idempotence and related requirements in edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 15(5), 484–491.
- Sander, P. T. (1988). *On Reliably Inferring Differential Structure from Three-Dimensional Images*. PhD thesis, Electrical Engineering, McGill University, Montreal, Canada.
- Sander, P. T., & Zucker, S. W. (1990). Inferring surface trace and differential structure from 3-D images. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 12(9), 833–854.

- Sebastian, T. B., Klein, P. N., & Kimia, B. B. (2004). Recognition of shapes by editing their shock graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 26(5), 550–571.
- Serra, J. (1982). *Image Analysis and Mathematical Morphology*, volume 1. Academic Press.
- Sethian, J. A. (1999). *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Cambridge Monograph on Applied and Computational Mathematics, 2nd edition. Cambridge University Press, UK.
- Siddiqi, K., & Kimia, B. (1995). Parts of visual form: Computational aspects. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 17(3), 239–251.
- Siddiqi, K., & Kimia, B. B. (1996). A shock grammar for recognition. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'96)*, (pp. 507–513).
- Siddiqi, K., Kimia, B. B., Tannenbaum, A., & Zucker, S. W. (2001). On the psychophysics of the shape triangle. *Vision Research*, 41(9), 1153–1178.
- Siddiqi, K., Kimia, B. B., Tannenbaum, A., & Zucker, S. W. (1999a). Shocks, shapes, and wiggles. *Image and Vision Computing*, 17(5–6), 365–373.
- Siddiqi, K., Shokoufandeh, A., Dickinson, S. J., & Zucker, S. W. (1999b). Shock graphs and shape matching. *International Journal of Computer Vision (IJVC)*, 35(1), 13–32.
- Sing, M., & Hoffman, D. D. (2001). Part-based representations of visual shape and implications for visual cognition. In T. F. Shipley & P. J. Kellman (Eds.), *From Fragments to Objects: Segmentation and Grouping in Vision*, Chapter 9 (pp. 401–459). Elsevier Science.
- Soille, P. (2004). *Morphological Image Analysis*, corrected 2nd edition. Springer.
- Stavroudis, O. N. (1972). *The Optics of Rays, Wavefronts, and Caustics*, volume 38 of *Pure and Applied Physics*. Academic Press.
- Sundar, H., Silver, D., Gagvani, N., & Dickinson, S. J. (2003). Skeleton based shape matching and retrieval. In *Proc. of IEEE International Conference on Shape Modeling and Applications* (pp. 130–142). Seoul, Korea.
- Tek, H., & Kimia, B. B. (2001). Boundary smoothing via symmetry transformns. *Journal of Mathematical Imaging and Vision*, 14(3), 211–223.
- Terzopoulos, D., Witkin, A., & Kass, M. (1987). Symmetry-seeking models for 3D object recognition. *International Journal of Computer Vision (IJCV)*, 1(3), 211–221.
- Thom, R. (1989). *Structural Stability and Morphogenesis*. Advanced book classics. Addison Wesley.
- Thompson, D. W. (1992). *On Growth and Form*. Dover Publications, complete revised edition of the 1917 original edition.
- Todd, J. T., & Reichel, F. D. (1989). Ordinal structure in the visual perception and cognition of smoothly curved surfaces. *Psychological Review*, 96(4), 643–657.
- Todd, S., & Latham, W. (1992). *Evolutionary Art and Computers*. Academic Press.
- Turing, A. (1952). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London*, 237(641), 37–72.
- van Tonder, G. J., & Ejima, Y. (2000). The patchwork engine: Image segmentation from image symmetries. *Neural Networks*, 13(3), 291–303.
- van Tonder, G. J., & Ejima, Y. (2003). Flexible computation of shape symmetries within the maximal disk paradigm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 33(3), 535–540.
- VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences*, 7(5), 207–213.
- Verrouast, A., & Lazarus, F. (2000). Extracting skeletal curves from 3D scattered data. *The Visual*

- Computer*, 16(1), 15–25.
- Wade, L., & Parent, R. E. (2002). Automated generation of control skeletons for use in animation. *The Visual Computer*, 18(2), 97–110.
- Weygaert, R. v. d. (2003). The cosmic foam: Stochastic geometry and spatial clustering across the universe. In E. Feigelson & G. Babu (Eds.), *Statistical Challenges in Modern Astronomy III* (pp. 175–196). Springer-Verlag.
- Wolfram, S. (1994). *Cellular Automata and Complexity: Collected Papers*. Addison-Wesley.
- Wolter, F.-E. (1985). *Cut Loci in Bordered and Unbordered Riemannian Manifolds*. PhD thesis, Technical University of Berlin, Department of Mathematics, Germany.
- Wolter, F.-E. (1992). Cut locus and medial axis in global shape interrogation and representation. Sea grant report, MIT, Cambridge, MA.
- Wolter, F.-E., & Frieze, K.-I. (2000). Local and global geometric methods for analysis interrogation, reconstruction, modification and design of shape. In *Proceedings of Computer Graphics International (CGI'00)* (pp. 137–151). Geneva, Switzerland. IEEE Computer Society.
- Wolter, F.-E., Peinecke, N., & Reuter, M. (2004). *Geometric Modeling of Complex Shapes and Engineering Artifacts*, volume 1 of *Encyclopedia of Computational Mechanics*, Chapter 16. Wiley.
- Wu, K., & Levine, M. D. (1997). 3-D shape approximation using parametric geons. *Image Vision and Computing*, 15(2), 143–158.
- Yuille, A., & Leyton, M. (1990). 3-D symmetry-curvature duality theorems. *Computer Vision, Graphics and Image Processing (CVGIP)*, 52, 124–140.
- Zucker, S. W. (1987). The emerging paradigm of computational vision. *Annual Review of Computer Science*, 2, 69–89.

Tracing axes of growth

Athanassios Economou

Introduction

One way of describing the structure of space, preferred by both Newton and Helmholtz, is through the notion of congruence. (Weyl 1954)

Hermann Weyl, one of the greatest mathematicians of the 20th century, in his famous lectures on symmetry in 1951, attempted to set up a formal framework for a theory of space pursued earlier by Leibniz, Newton and Helmholtz. The key idea in his method of describing the structure of space is the notion of congruence as a specific mapping that does not change the dimensions of bodies in space. The principle behind this idea is that the mapping can be extended for any point in space rendering thus notions of bodies and spaces as interchangeable; under this framework symmetries of bodies are equivalent to symmetries of space and symmetries of shapes are equivalent to symmetries of space. It is immaterial whether this mathematical framework can indeed provide an interpretative structure for the study of real space in general; the theory is significant as a study of pattern in general and in that sense provides a generous framework for new inquiries and insights for a variety of visual and aural patterns in design and the arts.

The idea of congruence is employed here for the discussion and representation of a specific subset of spaces that have one axis of growth. This specific class of designs is ubiquitous in architectural design and in the arts in general and yet a systematic effort of classifying and illustrating these structures for design purposes has yet to be undertaken. This can be partially explained because all classification schemes of symmetry structures require a mathematical sophistication not required or found any more in architectural or art curricula. Furthermore, the complexities of interactions of isometric transformations in three-dimensional space do not lend themselves easily to intuitive approaches to enumeration of classes of designs by simple trial and error efforts. Interactions of isometric transformations in two-dimensional space are relatively easy to comprehend and construct; for example, Alberti and Leonardo Da Vinci have been credited with a systematic clas-

sification and illustration of the cyclic and dihedral symmetries in the plane with their designs for roundel ornaments and centrally planned churches respectively (March 1998; Weyl 1954); alternatively, Frank Lloyd Wright and Le Corbusier have also been credited with a systematic but still not exhaustive illustration of structures with one or more axes of growth in the plane in their designs of housing complexes and urban design schemes (March & Steadman 1974). No architects or artists have been credited yet with a systematic construction of designs that take advantage of properties of growth along an axis in three-dimensional space.

This work looks closely at a specific class of three-dimensional designs that have one axis of growth and presents all possible algebraic structures that capture the symmetries of these designs. A specific set of designs is discussed, the symmmorphic designs, and is used as a framework to derive the non-symmmorphic designs and to complete all three-dimensional linear structures; the complete catalogue of all nineteen space structures that may be generated in this manner is presented in the end.

Linear growth

Spaces and spatial arrangements with one axis of growth are ubiquitous in nature and the arts. In the organic world this translatory symmetry, called by zoologists metamerism after the Greek word μεταμερισμός < meta-mer(os)-ism, Gk. meros = part, is quite frequent and it suggests a part-to-whole relation where the whole consists of transposed, translated parts. In architectural design this translatory symmetry appears in any configuration that consists of identical parts along an axis such as serial modular spaces, colonnades, friezes, entablatures, row housing, high-rise apartments and so forth. And many a contemporary architecture and artistic production characterized by relentless repetitions of spatial cells, including skins, tiles, modules, textures and so on, may indeed be adequately described by these structures.

In all these cases, these patterns may be small or large, simple or complex, coarse or smooth, discrete or continuous, abstract or concrete, or they may be presented in full semantic terms, for example walls, staircases, shelves, slabs, entablatures, row housing, street networks, and so forth. Still there is always a difference between mathematical abstraction and patterns of appearance; the former are abstract, infinite, geometrical, numerical; the latter are concrete, finite, corporeal, subjective. The same dualism exists here too; in mathematics any pattern that has translations extends in infinity, but in reality no spaces are infinite, and most of what we perceive as spaces or objects, at least for design purposes, is finite. It is the process of recognizing underlying transformations that is of interest here, rather than the actual finitude or infinitude of the pattern. For the purposes of

this work, and in general for any design consideration, a spatial pattern that exhibits translations is always finite. The number of translations that are necessary to characterize an arrangement is purely a matter of convention. For example, in archaeology, fragments of two copies with translatory symmetry are often enough to classify designs in terms of some underlying translational symmetry group (Washburn 1980). A nice account that attempts to eschew the problem of the number of sufficient translations to classify a pattern has been given within the field of shape grammars (Knight 1995).

The central question that rises here is whether such designs or indeed any other designs exemplifying one-dimensional growth take advantage of the structural ways of concatenating three-dimensional spaces or surfaces along an axis. Weyl in his lectures on symmetry asserted that the ancient Egyptians had explored all seventeen possible ways of unfolding pictorial ornamental designs on the plane (Weyl 1954); this is not to say the ancient artists knew that there were seventeen structurally possible ways of repeating a design in the plane. What Weyl suggests is the ancient artists had a very rich understanding of the medium and in their efforts to design these types of ornaments they exhausted all possible ways of combining a spatial module with copies of itself. It is very unlikely that a corresponding theme occurs in three-dimensional space. There is no panoramic or vantage point from which the designer or the researcher sees how the whole structure works and it is much harder to trace correspondences between spaces. It is the goal of this work to provide a lexicographic notation and a diagrammatic representation for each symmetry class and the spaces each denotes; it is hoped that this work will function as a map for the researcher or the artist and architect alike that want to explore these worlds.

Structures of space

A symmetry of a space is a transformation, such as a rotation, rotor reflection or any other isometric transformation, which leaves the space invariant; these transformations retain the shape and size of the space but alter its position. In three-dimensional space there are six isometric transformations, three direct and three indirect. The three types of direct isometries are rotations about a center, translations along a line, and screw rotations along a line; the latter are products of translations along a line with rotations about a center in the line. The three indirect transformations are isometries that alter additionally the handedness of a body of space and they include reflections along a plane, rotor reflections about a center, and glide reflections along a line; the rotor reflections are products of reflections with rotations and the glide reflections are products of reflections with translations. All six types of transformations are shown in Figure 1.

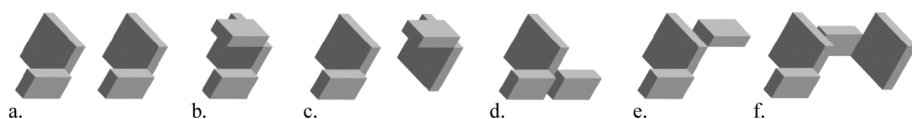


Figure 1. A pictorial example of all six generators in three-dimensional space. (a) Translation (b) Rotation (c) Screw rotation (d) Reflection (e) Rotor reflection (f) Glide reflection.

Information about the isometric transformations and the ways they are combined with each other is captured in the definition of the various symmetry groups. The mathematical study of the groups has been given in many sources (see for example, Armstrong 1988). A very brief account would suffice here; briefly, a group is a set endowed with a rule; the set can be any collection and the elements of the set are whatever comprises this collection. The rule combines any ordered pair x, y of elements of the set and obtains a unique product xy which also lies in the set; from this definition it follows that both possible ways of combining any two elements, x, y , that is, xy and yx , also lie in the set. The rule is usually referred to as a multiplication or a composition on the given set. A set G together with a rule on G becomes a group when it additionally satisfies three axioms: (a) The composition is associative, that is to say, $(xy)z = x(yz)$ for any three, not necessarily distinct elements in G . (b) There is an element e in G , called an *identity element*, such that $xe = x = ex$ for every $x \in G$. (c) Each element in G has an inverse x^{-1} which belongs to the set G and satisfies $x^{-1}x = e = xx^{-1}$. In general $xy \neq yx$; certain pairs of elements x, y in G obey $xy = yx$ and is said that these elements commute. The identity element e commutes with all elements of a group and every element commutes with its inverse. If all elements in G commute with each other, i.e., $xy = yx$ for all x, y of G , the group G is called commutative or Abelian. The order of the group is the number of elements in the set. A group is finite when the number of elements in the group is finite and infinite otherwise. A group is abstract if its elements are abstract, i.e., if they are not defined in any concrete way. A concrete example of an abstract group, i.e. a group with concrete elements with a law of composition, is called a realization of that abstract group. Such realizations might be groups of numbers, matrices, or geometric transformations.

The structure of a group is the statement of the results of all possible compositions of pairs of elements and it can be given either in analytical or a constructive way. The analytical description of the structure of the group is given usually in a square array, the multiplication table of the group; the constructive description is given in a set of group generators and defining relations that apply on the generators. This second approach describes groups independent of their order. For an element x in a group G , by the axiom on closure, it follows that all powers of x , that is, $x, xx, xxx, \dots$, all belong in the group. Furthermore by the axiom on inverses the elements $x^{-1}, x^{-1}x^{-1}, x^{-1}x^{-1}x^{-1}, \dots$ belong in the group. If all elements of the group

can be expressed as products involving only an element x and its inverse x^{-1} , then x is called the generator of the group. Similarly, for two elements x and y in a group G , then by the axiom on inverses, x^{-1} and y^{-1} are also in the group, and so are $x^{-1}yx$, $xyx^{-1}y$, and so on. Any product that can be written using x , y , x^{-1} and y^{-1} as factors in any sequence and with any finite frequency is an element of the group and is called a word (Baglivo & Graver 1976). If all elements of the group can be expressed as products involving two elements x and y and their inverses x^{-1} and y^{-1} , then x, y are called the generators of the group. The concept of group generators can be extended to a set of more than two elements. If S is a set of elements of a group G and if all elements of G can be expressed as products involving only the elements of S and their inverses, then the elements of S are the generators of G . The approach here in this work takes advantage of the definition and representation of a symmetry group entirely in terms of its generators.

A lot of the interest in groups and their corresponding designs lies in the nested structures observed within them. For any group G with an operation $(.)$, a subset H of G is a subgroup if H forms a group under the operation $(.)$. Not every subset of G is a subgroup; for example any subset of G that does not contain the identity element is bound not to be a group because it does not satisfy the group identity axiom. Still, subgroups are extraordinarily common; every group G is a subgroup of itself, and the identity element set $\{0^0\}$ is also a subgroup because it satisfies all group axioms. There is a very precise and concrete numerical relation between groups and subgroups: The order of a subgroup always divides the order of the group. This theorem, proved by Lagrange some time before even the abstract concept of a group was defined, shows that any group has two trivial subgroups, the identity and itself, and if it has more, these subgroups have to be in precise numerical relations to the group. This theorem plays a central role both in the construction of bigger groups in terms of smaller ones, the direct product groups, as well as in the division of bigger groups to smaller ones, the quotient groups; both techniques can be used for the enumeration of the complete list of the spatial patterns with one axis of growth (Armstrong 1988).

Among the most interesting realizations of abstract algebraic groups are the symmetry groups, i.e. the groups that contain isometric transformations. The symmetry groups can be additionally classified according to their translational structure and the dimensionality of the space that contains the transformations that belong in the set (Yale 1967). All symmetry groups are subgroups of the Euclidean group E consisting of all isometries, which is a subgroup of the similarities group S , consisting of all similarities. The ten symmetry groups E_{ij} in Euclidean space, for i = the number of axes of translation and j = dimension of space, are given in Table 1.

The complete enumeration of all the groups that comprise these ten structures has been carried out at different times, with different notations, and different

Table 1. Ten structures of space.

	0-Dimensional groups	1-Dimensional groups	2-Dimensional groups	3-Dimensional groups
Point groups	E_{00}	E_{01}	E_{02}	E_{03}
Line groups		E_{11}	E_{12}	E_{13}
Plane groups			E_{22}	E_{23}
Space groups				E_{33}

agendas in mind. Nice accounts in the literature are by Shubnikov and Koptsik (1974), and Toth (1964). In the notation offered here the subscripts ij provide an unambiguous way to discuss all these structures under a uniform labeling scheme. For example, within the convention of this scheme, E_{03} are the groups that capture the symmetry structures of three-dimensional spaces that have no translational structure, that is, spaces that have a center of symmetry; similarly E_{22} are the groups that capture the symmetry structures of two-dimensional spaces that have two distinct axes of translations, that is, essentially an infinity of emergent translations that fill out the plane. The structures of interest in this work are the E_{13} groups, that is, the groups that capture the symmetry structures of three-dimensional spaces that have one axis of translation, or alternatively, one axis of growth. These configurations appear in the literature with various names such as rods (Shubnikov & Koptsik 1974), fibers (Yale 1967), three-dimensional friezes and others.

The nineteen configurations with one axis of growth

The primary elements for the construction of the three-dimensional linear space patterns are the three-dimensional point space groups and the one-dimensional linear space groups. There are all together fourteen possible three-dimensional point groups that correspond to the fourteen types of spatial configurations around a center; the seven finite polyhedral groups of the platonic solids and the seven infinite prismatic groups. There are also three infinite linear space groups, the translation group, the screw rotation group, and the glide reflection group respectively (Armstrong 1988). Still, not all of these groups are candidates for the enumeration and this restriction applies primarily to the point groups. All polyhedral shapes such as cubes and the rest of the platonic solids and their derivatives, are shapes that have more than one primary axis of symmetry, and do not produce additional symmetrical configurations when they are combined with linear groups. This occurs because their axes of symmetry cannot be part of the emergent bigger structure; if they would they would pack space and they would create additional axes of growth. Similarly certain combinations of linear transformations,

that is, translations, rotor reflections and screw rotations along an axis do not necessarily produce always new structural possibilities and a careful strategy has to be followed.

The complete list of three-dimensional spaces with a singular axis can be generated from these seven point groups and the three linear groups in two ways that nicely correspond to the additive and the divisional mode of composition of groups. In the first mode the seven point groups are combined with the three one-dimensional linear space groups, that is, the translation group, the screw translation group and the glide reflection group to construct new direct product groups. Among the various combinations, those of the seven prismatic groups with simple translations along a line are quite special because they form seven basic linear configurations that retain the full symmetry of the pattern that is translated. These seven structures are called *symmorphic*, after the Greek term *συμμορφία* < *syn-morphe*, Gk. *morphe* = shape, because the whole configuration precisely retains the symmetry and form of the three-dimensional module that is repeated along the line. In fact the total number of the *symmorphic* groups is nine because there are two more that are characterized from half-turns embedded in the translational axis; still, both these types are limiting cases for the other seven *symmorphic* types and they will not be dealt here separately. The rest of the designs that have a translational structure can be generated by the successive combination of the seven infinite prismatic groups with the other two linear space groups after carefully eliminating emergent duplicating cases. A nice account of this method is found in Shubnikov and Koptsik (1974).

In the second mode the seven *symmorphic* groups are dissected to produce symmetry subgroups; this is the approach adopted in this work. An interesting aspect of this approach is that the emergent module that appears to be translated along a line is always composed by smaller parts whose relationship to one another is evident only in the formation of the bigger pattern and not in the description of the module itself. These patterns have the characteristic that the translated module appears to have no symmetry in itself, but that the whole configuration has additional symmetries to its translational structure. This is a straightforward result from the fact that the symmetries of the original module are dropped but not the parts themselves; as a result these parts are located in a new position and the emergent symmetry structure of the pattern becomes evident only after a sufficient number of translations of this version of the module. A complete catalogue of all linear three-dimensional symmetry classes is given below in seven sets that are headed by an initial *symmorphic* structure and followed by all possible non-*symmorphic* structures that can be extracted from that. The notation adopted here is the so-called non-coordinate notation for symmetry classes (Shubnikov & Koptsik 1974). This notation indicates the number and types of symmetry generators and their corresponding spatial relationships. More specif-



Figure 2. Spatial patterns produced by three generators: (a) 4. (b) 4: m . (c) 4: $m.a$.

ically the symmetry symbols are: a rotation axis n of order n , a mirror-rotation axis ($2\bar{n}$) of order $2\bar{n}$, a reflection plane m , a translation axis a with an elementary translation a , and a glide reflection axis ($\bar{a}$) with an elementary translation $a/2$. The signs between the symmetry elements denote the spatial relationships between the symmetry elements. The two-point sign ($:$) between two elements indicates that these symmetry elements are perpendicular to one another; the one-point sign ($.$) indicates that these symmetry elements are parallel to one another; lastly, an oblique-stroke ($/$) sign indicates that these symmetry elements are inclined to one another; still this last sign is not employed for the classes of design discussed here because an oblique axis to an axis of translation would generate a second axis of translation and therefore an infinite number of translations that fill the plane. These symbols are enough to describe the symmetry of any spatial configuration in three-dimensional space. For example, a pattern notated as 4 has one generator, the four-fold axis of rotation, and consists of four symmetries; a pattern notated as 4: m has two generators, a four-fold axis of rotation and a mirror reflection plane m perpendicular to the axis of rotation, and consists of eight symmetries; a pattern notated as $a.4:m$ has three generators, an axis of translation a parallel to a four-fold axis of rotation and perpendicular to mirror reflection planes m , and consists of infinite symmetries. Figure 2 shows the three spatial patterns discussed; the translational pattern is shown for a single application of a translation.

There are various ways of representing symmetry transformations; for brevity and diagrammatic clarity here all symmetry classes are pictorially represented in an orthographic projection to render unambiguously the results of the applications of the transformations. The number of labels denote the number of identical parts in a configuration; dashed labels denote the original position of a label. Black labels denote the position of the label towards the viewer and white towards the back. Numbers denote the internal repetitions of a label along the translational axis within the duration of a complete translation of the pattern. All axes of transformations are assumed to be perpendicular to the axis of viewing. Figure 3 illustrates a diagrammatic representation of the six transformations of three-dimensional spaces.

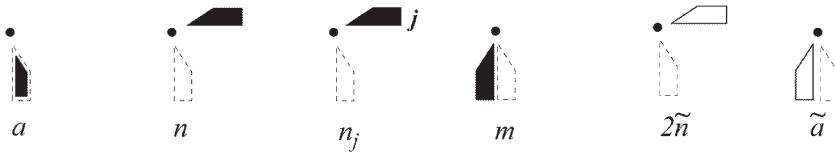


Figure 3. A diagrammatic example of all six generators in three-dimensional space. (a) Translation (b) Rotation (c) Screw Rotation (d) Reflection (e) Rotor reflection (f) Glide reflection.

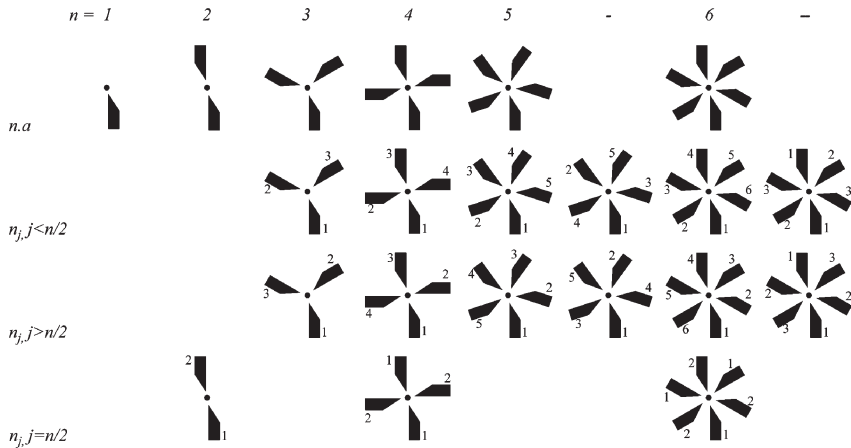


Figure 4. The four types of linear spaces based on the class $n.a$, for $n \leq 6$.

$n.a$

The space $n.a$ is generated by successive translations of shapes with symmetry n along their primary axis of rotation at a distance a . The combinations of the rotations n and the translations a , produce the screw rotations n_j , for $j < n$. A substitution of the symmetry axis n by the screw axis n_j produces three types of subgroups n_j , for $(j < n/2)$, $(j = n/2)$, $(j > n/2)$. The $j < n/2$ and $j > n/2$ classes are isomorphic; the first denotes a clockwise configuration, the latter a counter-clockwise one. The $j = n/2$ is a neutral one in the sense that is simultaneously left-handed and right-handed. The four possible groups in this category are the $n.a$, and n_j for $(j < n/2)$, $(j = n/2)$, $(j > n/2)$. A diagrammatic representation of these four types of patterns, for $n \leq 6$, is given in Figure 4.

$2\tilde{n}.a$

The space $2\tilde{n}.a$ is produced by successive translations of shapes with symmetry $2\tilde{n}$ along their rotor reflection axis at a distance a . By definition this symmetry class occurs only for shapes that have a mirror-rotation axis of an even order $2\tilde{n}$, that

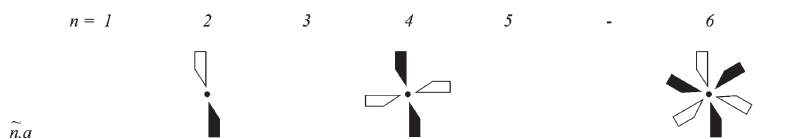


Figure 5. The single type of a linear space based on the class $[2\tilde{n}.a]$, for $n \leq 3$.

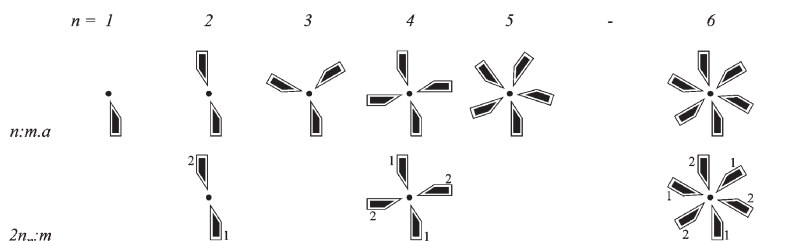


Figure 6. The two types of linear spaces based on the class $n:m.a$, for $n \leq 6$.

is, $\tilde{2}$, $\tilde{4}$, $\tilde{6}$, and so forth. There are no allowable substitutions in this category. A diagrammatic representation of the pattern $2\tilde{n}.a$, for $n \leq 3$, is given in Figure 5.

$n:m.a$

The space $n:m.a$ is produced by successive translations of shapes with symmetry $n:m$ along their primary axis of rotation at a distance a . The combinations of translations and mirror planes produces new reflections at distances a as well as at their midpoints at $a/2$. Other emergent symmetries include screw rotations and rotor reflections. A substitution of the symmetry axis n by the screw axis n_j produces one more type of a subgroup $n_j:m$ for ($j = n/2$) or alternatively $2n_n:m$. The other two possibilities for ($j < n/2$) and ($j > n/2$) do not produce new types because the mirror planes are not part of the overall pattern. The two possible groups in this category are the $n:m.a$ and $2n_n:m$. A diagrammatic representation of the two types of patterns, for $n \leq 6$, is given in Figure 6.

$n:2.a$

The space $n:2.a$ is produced by successive translations of shapes with symmetry $n:2$ along their primary axis of rotation at a distance a . The products of the rotations n with the half-rotations produce new emergent half-rotations at distances a as well as at their midpoints at $a/2$. The combinations of the rotations n and the translations a , produce the screw rotations n_j . Similarly to the first case of these symmorph spaces, a substitution of the symmetry axis n by the screw axis n_j produces three types of subgroups $n_j:2$, for ($j < n/2$), ($j = n/2$), ($j > n/2$). The four possible groups in this category are the $n:2.a$ and $n_j:2$ for ($j < n/2$),

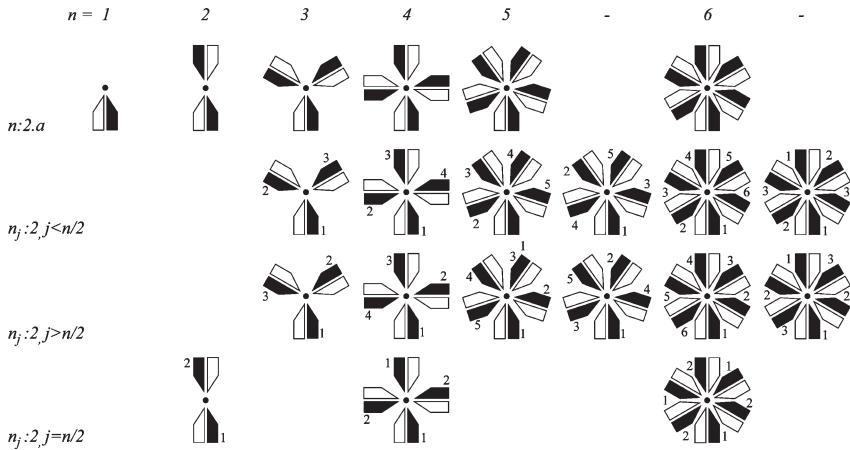


Figure 7. The four types of linear spaces based on the class $n:2.a$, for $n \leq 6$.

($j = n/2$), ($j > n/2$). A diagrammatic representation of these four types of patterns, for $n \leq 6$, is given in Figure 7.

$n.m.a$

The space $n.m.a$ is produced by successive translations of shapes with symmetry $n.m$ along their primary axis of rotation at a distance a . The products of the rotations n with the mirror plane produce new emergent mirror bisecting the angle of the rotation n , for n even number of rotations. Other emergent symmetries include screw rotations n_j , and glide reflections $\tilde{a}$. A substitution of the symmetry axis n by the screw axis n_j produces the space $n_j.m$ for ($j = n/2$) or alternatively $2n_n.m$. The other two possibilities for ($j < n/2$) and ($j > n/2$) do not produce new types because the mirror planes are not part of the overall pattern. A substitution of the mirror plane m by a glide reflection $\tilde{a}$ produces the space $n.\tilde{a}$. The three possible groups in this category are the $n.m.a$, $n.\tilde{a}$ and $n_j.m$, for ($j = n/2$). A diagrammatic representation of these three types of patterns, for $n \leq 6$, is given in Figure 8.

$2\tilde{n}.m.a$

The space $2\tilde{n}.m.a$ is produced by successive translations of shapes with symmetry $2\tilde{n}.m$ along their axis of rotor reflection at a distance a . The products of the rotor reflections $2\tilde{n}$ with the mirror planes m produce new emergent half-turn rotations bisecting the angle of the mirror planes at distances a as well as at their midpoints $a/2$. Other emergent symmetries include glide reflections $\tilde{a}$ along the mirror planes at distances a . A substitution of the mirror plane m by a glide reflection $\tilde{a}$ produces the space $2\tilde{n}.\tilde{a}$. The two spaces in this class are the $2\tilde{n}.m.a$ and

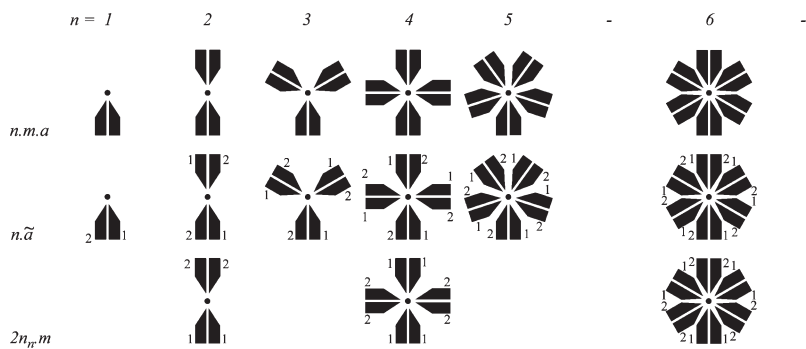


Figure 8. The three types of linear spaces based on the class $n.m.a$, for $n \leq 6$.

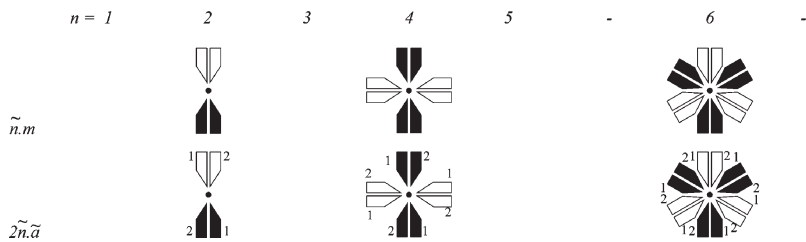


Figure 9. The two types of linear spaces based on the class $2\tilde{n}.m.a$ for $n \leq 3$.

$2\tilde{n}.\tilde{a}$. A diagrammatic representation of these two types of patterns, for $n \leq 3$, is given in Figure 9.

$n.m:m.a$

The space $n.m:m.a$ is produced by successive translations of shapes with symmetry $n.m:m$ along the axis of rotation at a distance a . The emergent symmetries are many here because of the presence of the four generators and they include screw rotations n_j , glide reflections $\tilde{a}$, reflections m bisecting the angle of rotation n , reflections m along perpendicular mirror planes at distances a as well as at their midpoints at $a/2$. A substitution of the symmetry axis n by the screw axis n_j produces the space $n_j.m:m$ for ($j = n/2$) or alternatively $2n_r.m:m$. The other two possibilities for ($j < n/2$) and ($j > n/2$) do not produce new types because the mirror planes are not part of the overall pattern. A substitution of the mirror plane m perpendicular to the axis of rotation n by a glide reflection $\tilde{a}$ produces the space $n.\tilde{a}:m$. The three spaces in this class are $n.m:m.a$, $2n_r.m:m$, and $n.\tilde{a}:m$. A diagrammatic representation of these three types of patterns, for $n \leq 6$, is given in Figure 10.

The complete list of all possible three-dimensional configurations that have an axis of growth consists of nineteen classes. A comparative illustration for all

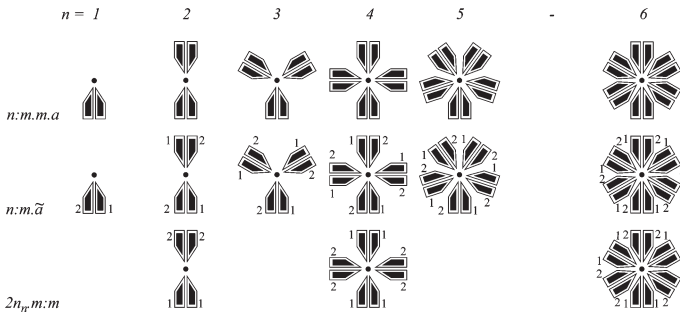


Figure 10. The three types of linear spaces based on the class $n:m:m.a$, for $n \leq 6$.

these classes is given in Figure 11. All nineteen types of spatial patterns are illustrated in terms of a simple spatial relationship between a square prism and an oblong; seven of these patterns, the symmorph configurations, are comprised by translated copies of the spatial relation of the square prism with the rectangular prisms. The other twelve, the non-symmorph configurations, are comprised by internally translated parts of this spatial relation; eight of these twelve configurations are comprised by translated copies of two transposed parts, the one in front of the other along the axis of growth; these are the patterns that have mirror planes in their structure. The other four of the twelve are essentially two pairs, two clockwise and two counterclockwise spaces and these are the patterns that have only screw rotations in their structure. The catalogue is structured around the seven symmorph classes to foreground the nested relationships of groups within groups.

Discussion

One of the most beautiful aspects of group theory and pattern making is the nested relationship of groups within groups. This relation has a strong aesthetic significance in works of art because several part-to-whole relations that are discerned in art and architecture bear a strict relation to this nested order of structures of space. In general, the construction of abstract groups as aggregations of smaller groups or as divided parts of bigger groups corresponds to basic modes of composition in architecture and in the arts. In the first case, the so-called additive mode of composition, shapes are combined in some way to produce new shapes; in the second case, the so-called divisional mode of composition, shapes are divided to produce new parts. Classic examples of both modes can be found in Palladio's and Alberti's work respectively; the former in Palladio's work and his assembly of rooms and ratios of 1:1, 1:2, 2:3, 3:4, and the latter in Alberti's work in his

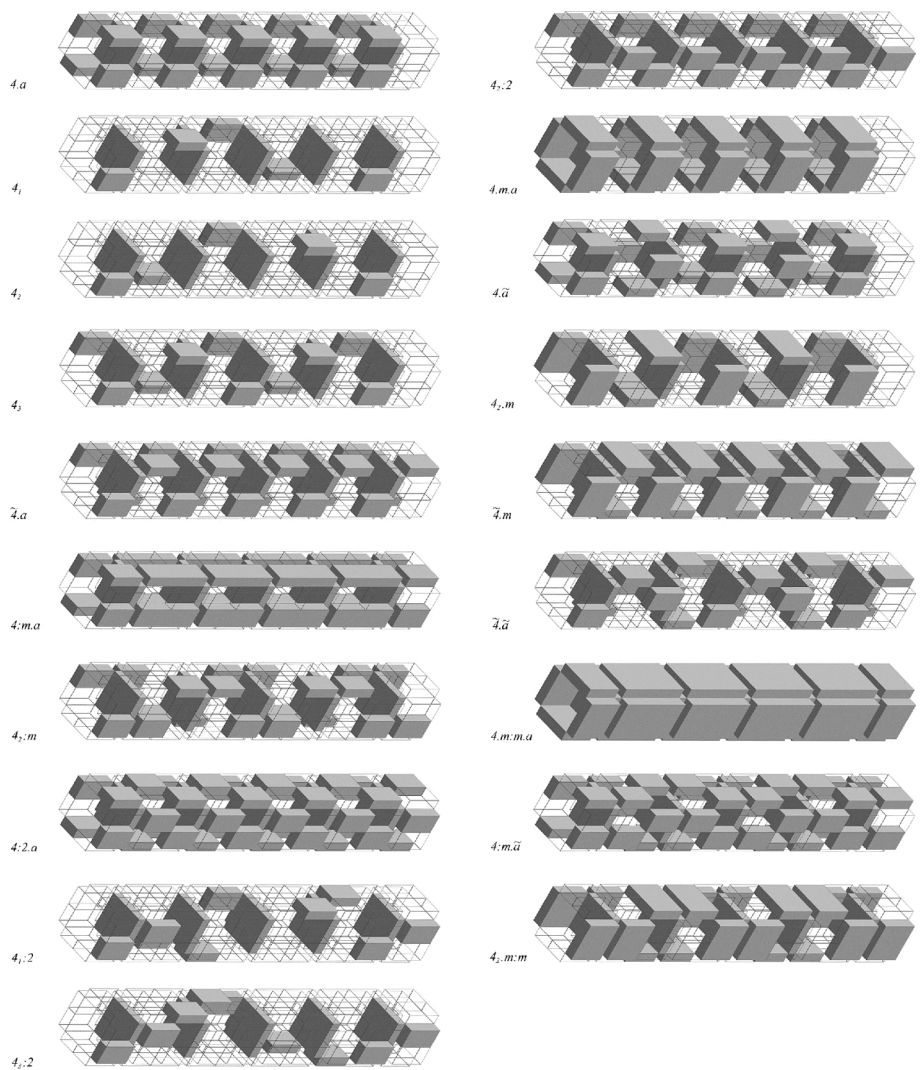


Figure 11. The nineteen possible spaces based on a simple spatial relation between a prism and an oblong.

subdivisions of the Doric column in proportional relationships (1:3:9) (March & Steadman 1974). Both approaches and worldmakings provide alternative descriptions of patterns, abstract or concrete, mathematical or aesthetic, and they both illuminate the notion of constructed nested hierarchies in pattern making.

A major motivation underlying this work is the exploration of the relationship between languages and configurations. Languages of design are generated by all different kinds of rule-based systems including shape grammars, cellular au-

tomata, L-systems and so on (see for example, Stiny 1976); configurations are similarly explored to discern what is structurally possible in a design context (March 1998; Economou 1999). Both provide a complementary insight in design explorations and both suggest alternative ways to tackle and unravel complexity in design. The work here privileged the second mode of design and explored a specific configurational problem, the class of three-dimensional spaces characterized by an axis of growth. The nineteen types of group structures that capture the properties of these spaces were briefly described and illustrated to show pictorially the expressive power of the structures of these spaces.

References

- Armstrong, A. (1988). *Groups and Symmetry*. New York: Springer-Verlag.
- Baglivo J, & Graver, J. (1976). *Incidence and Symmetry in Design and Architecture*. Cambridge: Cambridge University Press.
- Economou, A. (1999). The symmetry lessons from Froebel's building gifts. *Environment and Planning B: Planning and Design*, 26, 75–90.
- Knight, T. (1995). Constructive symmetry. *Environment and Planning B: Planning and Design*, 2, 419–450.
- March, L., & Steadman, P. (1974). *The Geometry of Environment: An Introduction to Spatial Organization in Design*. Cambridge, MA: MIT Press.
- March, L. (1998). *Architectonics of Humanism*. London: Academy Editions.
- Shubnikov, A., & Koptsik, V. (1974). *Symmetry in Science and Art*. Trans. D. Harker. New York: Plenum Press.
- Stiny, G. (1976). Two exercises in formal composition. *Environment and Planning B*, 3, 187–210.
- Toth, F. (1964). *Regular Figures*. Oxford: Pergamon Press.
- Washburn, D., & Crowe, D. (1988). *Symmetries of Culture: Theory and Practice of Plane Pattern Analysis*. Seattle: University of Washington Press.
- Weyl, H. (1954). *Symmetry*. New Jersey: Princeton University Press.
- Yale, P. (1968). *Geometry and Symmetry*. New York: Dover Publications.

John Willats†

It was a sad moment when I received Ruth Willats' mail about John's recent death. I am not quite sure when I met John for the first time, but I think he probably approached me because of my 1982 paper (with Andrea van Doorn) "*The shape of smooth objects and the way contours end*". John was much interested in the topic, he mailed me his study on the depiction of smooth forms in paintings by Paul Klee. John had worked on the general topic for quite a time already, as I knew from his wonderful little book "*Drawing Systems*" with Fred Dubery of 1972.

In 1987 I happened to be Douglas Holder Fellow of University College and spent a period at Oxford. I took the opportunity to visit John who then lived at Bradford on Avon (Wiltshire, UK). A small town of 10 000 inhabitants, it lies on the Vale of Pewsey, partly on the hill. John lived in a small house in town with his first wife. We walked along the Kennet and Avon Canal (which – quite spectacularly – runs high over the valley on an aqueduct). The canal was completed 1810 and runs 57 miles long between the River Thames (at Reading) and the River Avon (at Bath). Before the railways (mid 1800's) it carried barges between Bristol and London. John was much interested in the history and we discussed canal technology over pints of beer in local pubs. At that time John was restoring an old cottage at the edge of town. The building needed quite a bit of work as the stone was crumbling in many places. John had set up a small workshop in which he could saw the Jurassic limestone ("Bath Stone" from the Mendips) from which the older buildings are constructed over there. He taught me how you should always use the stone in the same orientation as it was quarried from the mountain in order to prevent premature weathering. I remember spending an afternoon carrying rough cut flat stones to the roof on which they were hung on bronze pegs, smaller stones high up, larger ones lower down, in the old style. It was hard labour, but John (who must have been doing this on a daily basis) remained cheerful as ever.

Tragically, John's first wife died early, and I don't think he actually lived in the cottage very long. After he remarried he lived with his partner in Somerton Somerset. Although John invited me on many occasions, I never had the opportunity to visit them there. Apparently John still kept his ties with his former community, for in 2000 the Mayor of Bradford on Avon unveiled "Millie" which must be one of John's larger sculptures. He produced it in response to a national competition organised by Bradford on Avon's Town Council. The re-design of the Festival Garden

site, in which the statue stands, was also part of the competition. “Millie” refers to the mill girls from Bradford on Avon’s wool and cloth trade. The sculpture is a 14 feet high bronze cast. John made a number of larger sculptures that I haven’t seen although John at times mailed me pictures of them. Most are very interesting “flat sculptures” that John hammered out of plate. Even when his health was already failing him, John explained to me how he would prepare a concrete mould in which he would hammer the metal plates into shape before welding them together. These must have been major undertakings and can hardly have been silent enterprises.

Most pieces I have seen of John’s work were smaller objects of the tin smith craft, galvanized for outdoor service. They were mostly amphora or flower shapes that John sold as garden decorations. They were really very charming and I regret not having invested in some of them. The larger works, that I know only from pictures, evidently clearly reflect John’s philosophy on visual perception and artistic production. We discussed these on various occasions because of our common interest in vision and especially pictorial shape and space.

As I grew older myself I came in a position to invite John to conferences and workshops or recommend or suggest him to other organizers. These were then the only opportunities where we met personally, although we remained in correspondence over topics of common interest. These were mainly pictorial space and shape, but also artistic expression.

John Willats read engineering at the University of Cambridge (where he was awarded a prize for his engineering drawings) before he trained as a sculptor at the Royal College of Art (where he gained the College drawing prize). This probably explains his interest in science and such things as computer rendering. In his book with Fred Dubery there are a number of computer drawings, based on measurements John did on a Greek sixth-century stone carving at the British museum. John used a surveyor’s digitizer and prepared the drawings on an IBM1130 (introduced 1965, CPU with 8K words of 280 kHz core memory, John would have used FORTRAN) with a graph plotter at North-East London Polytechnic. The digitizing must have been a major job. John mailed me the (barely readable, typewritten) data in the 1990’s. It was hard to get the numbers out of the typewritten text, it being often unclear where one number ended and the next began. With difficulty we produced a triangulation that enables modern computer graphics. I generated a few for some of John’s papers and I frequently use the data myself to prepare various examples. It has become a memorial.

John understood at the gut level that *vision* happens to you, like sneezing, or as you might say is “presentation”. Of course there is also visual *cognition*, which is a kind of thought for which you may be held responsible, unlike the presentations for which you cannot be held responsible. Although I really appreciated John’s grasp of what vision is like, I never quite understood his frequent references

to Marr's work, which to me reflect a completely different grasp of the ontology of vision. However that may be, John and I had many tangencies on the occasions we were able to compare notes. In meetings John was refreshingly outspoken and I often enjoyed his mingling in discussions. John had a characteristic clear voice and spoke the perfect Queen's English that makes us continental scientists feel perhaps slightly inferior or certainly at a severe disadvantage. He had very strong convictions on the topic of art education (and, on the whole convictions I share: imagine what would happen to math education when you would keep students from contacts with the past in order not to stifle their natural creativity?) and on one occasion I still remember well, he stood up to present his views with some force for a full five minutes or so, after a talk by the wife of the organizer who (coincidentally) happened to be an art teacher. One time, on the topic of the importance of context in visual recognition John recalled an amusing incident from the time he attended art school. As he rode on a bus home a girl, not without considerable charms, addressed him as if they met daily although John did not have the slightest idea who she might be. They kept on the conversation on a number of general topics till the girl got off the bus. John was puzzled. It was only a considerable time later when it suddenly occurred to him that he actually spent at least an hour a day looking intently at the girl in art class. She was a professional model and he had never before seen her with her clothes on.

John's scientific work can be appreciated best from his wonderful books "*Art and Representation*" and "*Making Sense of Children's Drawings*". In "*Art and Representation*" John presents his theory of pictures. Most importantly, he discusses a vocabulary that enables a scientific discussion of the topic. In "*Making Sense of Children's Drawings*" John applies his insights to the study of children's drawings. This is one side of John's activities that I never witnessed directly, he would visit schools and collect data on the drawing of children of various age groups. His views are, as was to be expected, refreshing and make short work of existing (indeed predominantly mythical) views on the matter.

Last time I met John was at a meeting at the Mitteleuropa Foundation at Bolzano, organized by Liliana Albertazzi. John was not very well then, bodily I mean, for his mind was keen as ever. We planned some future collaboration in which John was going to apply the tools we (Andrea van Doorn and myself) had forged to enable quantitative psychophysics of pictorial objects to his children studies. He mentioned that he planned to visit Venice after the meeting in order to look at some stone lions at San Marco. He would like to sculpt a stone lion in Romanesque style for his wife, for her to see when working in her study. I pointed him to some remarkable instances of such stone lions at Bolzano itself. John later emailed me that he indeed went to see them, even though he at the time walked slowly and only with difficulty. It turned out to be the last I would hear of him.

John leaves his wife Ruth and son Bill to which he dedicated his last book “*Making Sense of Children’s Drawings*”.

Jan Koenderink

References

- Dubery, F., & Willats, J. (1972). *Drawing Systems*. London: Studio Vista.
- Koenderink, J. J., & Van Doorn, A. (1982). The shape of smooth objects and the way contours end. *Perception*, 11, 129–137.
- Willats, J. (1980). On the depiction of smooth forms in a group of paintings by Paul Klee. *Leonardo*, 13(4), 276–282.
- Willats, J. (1997). *Art and Representation*. Princeton University Press.
- Willats, J. (2005). *Making Sense of Children’s Drawings*. NY: Lawrence Erlbaum Associates.



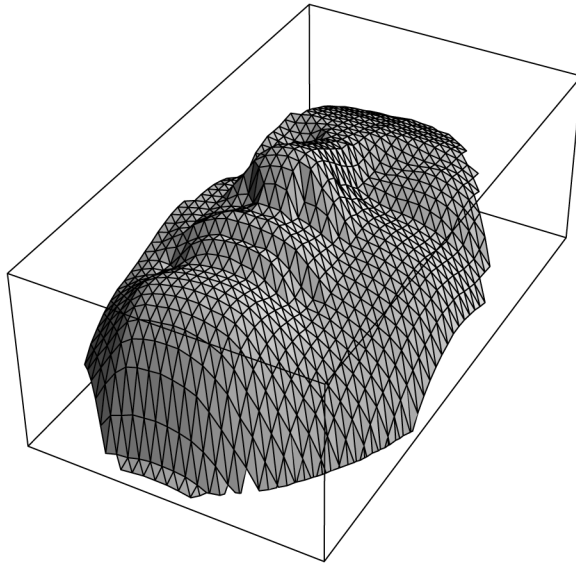
“Millie” by John Willats – Bradford on Avon. See: <http://www.pmsa.org.uk/gallery/cf.htm> and <http://www.pmsa.org.uk/gallery/windows/boa-text.htm>.



Dundas aqueduct, the Kennet and Avon Canal



The Old Bridge Teashop (small building in the middle) Bradford on Avon



Sixth-century stone carving, digitized by John Willats

Name index

A

Adelson, E. H. 43–48, 50, 52,
55
Akino, A. 278
Akisato, R. 273–276
Albertazzi, L. 66, 69, 161
Alberti, L. B. 225, 351, 364
Alexander, C. H. 259
Alpers, S. 230
Archimedes 317
Arnheim, R. 12, 20, 23, 72,
74–75, 81, 86–90, 93, 166,
171, 176–178, 184–85, 189,
195, 197, 208–209, 216, 235,
251, 258–259, 261–263, 269,
276, 278, 281
Asada, H. 323

B

Bach, J. S. 289
Baldassarre Castiglione 231
Balla, G. 187
Ballard, H. D. 253
Barsalou, L. W. 254
Bayes, Th. 139
Benussi, V. 17, 99, 169, 175, 181
Bergevin, R. 334
Berkeley, G. 138–139, 141
Beverwyck, J. van 227
Bharucha, J. J. 79
Biederman, I. 334
Blessum, J. A. 90
Blum, H. 318, 323, 327, 329
Boccioni, U. 26, 186–187
Bolyai, J. 319
Brady, M. 323
Brancusi, C. 142, 145
Brandt, S. A. 251
Braque, G. 195, 209
Brentano, F. 16, 27, 166–168,
176
Bruner, J. 161

Brunswick, W. 17
Bucher, N. M. 73

C

Canny, J. 129
Cartan, E. 120
Cauchy, A. L. 314
Cézanne, P. 26
Chafe, W. 241
Cheng, J. L. 79
Choisy, A. 211
Chomsky, N. 211
Cizek, F. 209
Clark, A. B. 197
Clark, K. Lord 225
Clowes, M. B. 211

D

Dawkins, R. 335
Debussy, Ch. 289
de Hollanda, F. 224
Delaunay, R. 186
De Noyers, S. de 224
Depero, F. 189
Dickens, C. 289
Drebbel, C. 225–226
Dubery, F. 367–368
Duchamp, M. 186
Dudreville, L. 187
Duncker, K. 167
Dupin, C. 305
Dürer, A. 225–226, 229–230

E

Edelman, S. 112
Ehrenfels, Ch. von 17
Euclid 230–231
Eyck, J. van 225

F

Fechner, Th. 168, 178
Fermat, P. de 316

Fiedler, K. 140
Fineberg, J. 212
Finke, R. A. 76–77, 251
Fodor, J. J. 15
Fourier, Ch. 308
Freyd, J. J. 76–77, 79–81, 87,
90
Fry, R. 209

G

Georgopoulos, A. P. 109
Giacometti, A. 187
Gibson, J. J. 196, 221–222, 228,
232, 320
Giblin, P. J. 330
Goldmeier, E. 305
Goodman, A. W. 316
Gris, J. 211
Gombrich, E. 137, 221–222,
232

H

Hamilton, W. R. H. 309
Hartmann, N. 24, 166
Hawkwood, J. 231–232
Hebb, D. O. 253
Hedberg, B. 251
Heisenberg, W. K. 306
Helmholtz, H. von 221, 224,
226, 228, 351
Herbart, J. F. 167
Hering, E. 224
Hildebrand, A. von 140–141,
144–145, 149–150, 189
Hochberg, J. 232
Hockney, D. 232
Hoffman, D. D. 313, 333, 338
Holsanova, J. 238, 251
Hoogstraten, S. van 226
Hough, D. 308
Hubbard, T. 79, 90
Hubel, D. 119, 340

Huffman, D. A. 211
Husserl, E. 169, 178, 185
Huygens, C. 225, 315

I

Ingarden, R. 24

J

Jacobi, C. G. 313
Johansson, G. 332
Johnson, J. Q. 90
Johnson, K. 109

K

Kandinsky, W. 166, 168, 178,
180, 187, 195, 209, 212
Kanizsa, G. 19–21, 168
Kenkel, F. 174, 176
Kennedy, J. 206
Kepler, J. 225, 228, 230
Kimia, B. B. 318, 330, 339
Klee, P. 166–168, 171, 173, 176,
179, 184, 196, 199, 209,
211–213, 215–217, 367
Klein, F. 141, 318
Koch, H. von 321
Koenderink, J. J. 304, 306
Koffka, K. 12, 17
Köhler, W. 13, 15–18
Koptsik, V. 356–357
Kosslyn, S. M. 253
Kovács, I. 332, 340

L

Laeng, B. 251, 253
Lagarde, J. 310
Lagrange, J.-L. 309, 355
Langer, M. S. 331
Le Corbusier (Ch.-E.
Jeanneret) 352
Leonardo da Vinci 144, 351
Léger, F. 209
Leibniz, G. 351
Lévy, P. 321
Levi-Civita, T. 323
Levine, M. D. 334, 339
Leymarie Fol, F. 330, 339
Leyton, M. 66, 72, 73, 87,
91–92, 182, 307, 322, 329,
333, 335–339
Lipps, Th. 13, 178, 180
Lipschitz, R. O. S. 317

Lobachewski, N. I. 319
Lord, E. A. 318
Lorentz, H. A. 312
Luquet, J.-H. 206, 210

M

Mach, E. 224, 305
Magritte, R. 148
Mandelbrot, B. 321
Marc, F. 209
Marees, H. von 140
Marey, E.-J. 332
Marr, D. 36, 37, 38, 47, 50, 195,
197–198, 204, 369
Masin, G. C. 103
Massironi, M. 181–182
Mast, F. W. 253
Matisse, H. 195, 209
Meinong, A. 10
Metzger, W. 21, 102, 179–180,
184
Michelangelo 224, 232
Michotte, A. von 17–18, 169,
179
Minkowski, E. 312, 315, 319
Monge, G. 309
Montanari, U. 318
Mozart, W. A. 289
Musatti, C. 21, 99, 175
Musgrave, K. 321
Münter, G. 212

N

Nagai, M. 87
Nakane, K. 274
Newton, I. 351
Nicholls, A. L. 206
Nilsson, N. 251
Nishihara, H. K. 197
Nordqvist, S. 251
Norman, J. F. 318, 319, 320
Notton, D. 247

P

Palladio 60
Palmer, S. E. 73
Panofsky, E. 225
Panzer, T. M. 79
Pentland, A. P. 43–48, 50, 52,
55, 334
Picasso, P. 26, 168, 195, 209
Plato 18

Poincaré, H. 312
Poli, R. 18
Poussin, N. 224–225
Protagoras 224
Ptolomy 230–231
Pylyshyn, Z. 15, 253

R

Raphael 230–231
Reichel, F. D. 320
Rensink, R. A. 9
Richards, W. 333, 338
Richardson, M. 209
Ritoh, A. 273
Rivière, J. 209–210

S

Sander, P. T. 324, 329
Schiller, F. 221
Schoenberg, A. 289
Shigemori, M. 261
Shimoyama, S. 261
Shingen 280
Shubnicov, A. 356
Simon, H. A. 288–290,
292–296, 298
Stark, L. W. 247, 251
Steadman, P. 226–227
Stroll, A. 36, 66
Strubecker, K. 149

T

Tek, H. 339
Teodorescu, D.-S. 251, 253
Thom, R. 340
Thompson, D. W. 72, 321
Titchener, E. B. 151
Todd, J. T. 318–320
Tonder, G. van 273, 297
Toth, F. 356
Turing, A. 333

U

Uccello, P. 232
Ulam, S. 334

V

Varela, F. J. 181
Vermeer, J. 225–227

W

Wertheimer, M. 178–179
Weyl, H. 351, 353
Wiesel, T. 119, 340
Willats, J. 185, 206, 368
Wilson, C. B. 318
Wolfram, S. 335
Wolter, F.-E. 328

Wright, L. F. 352

Wu, K. 334

Y

Yagi, A. 87

Yaglom, I. M. 149

Yarbus, A. L. 247

Z

Zanette, D. H. 290

Zeki, S. 188

Zeno 4

Zipfs, G. K. 288–296, 298

Zouen 259, 280, 285

Zucker, S. 322, 324, 329, 331

Subject index*

- A**
Accommodation 46–56
Act 3
 Deployment 170
 Intentional 16, 188
Affordance 177, 221, 223
Architecture 60–61, 65
 Classical 60
Attention 64, 253
 Deployment of 65, 69
 Shift of 253
Axis
 Medial 266, 268
 Of growth 356
 Symmetry 336, 356
- B**
Bayesian approach 43, 50, 66, 68
Belongingness 8
Bilateral symmetry 262–263
Binocular disparity 36–37, 47, 49, 57
Biomorphs 355
Bonsai 287
Boundary 20
 Flow 127
 Enhancement 60, 63, 64
 of shading flow 116, 127
Brightness 38, 46, 55, 67
- C**
Camera obscura 196, 225
Camouflage 49
Catastrophe Theory 340
Category
 Emergent 21
 Level 18
 Universal 18
Change 8–11
 Brightness 54–55
 Geometric 50
 in a spatiotemporal structure 9
 Lightness 38, 53
 Reflectance 38, 53, 60
 Structural 50
Chronophotographs 332
Co-circularity 119, 322, 328
Cognitive maps 157
Compatibility function 121
Computer interfaces 63
Completion 26, 166, 180–181, 297
Conservation laws 314
Constitutive assumptions 314
Constraints 40, 46
Contact circle 327
Content 23–25, 173
Continuity 37, 67
Contour
 Fragments of 108
 Illusory 223
 Orientation 106
Convergence 47, 68
Coplanar lightness change 53–55
Corporeal schema 11
Cost function 44
Critical net 330
Cue 36, 57, 66, 68–69
 Ambiguity 146
 Combination 68
 Conflict 47, 53, 57, 69
 Depth 57, 138
 Integration 44
 Monocular 47, 68
 Pictorial 138, 141
 Primary 139
 Secondary 139
Curvature 36, 118
 Discontinuity 56
 Normal 120
 Tangential 120
Curve 118
Cut 127
 Cut locus 328
CW-complex 331, 343
- D**
Deformation 175, 183–184
 of a continuum 313
 General 313
Depiction 26
Depth
 Pictorial 139–140
 Visual 138, 272
Design 40, 57–58, 63, 65, 260, 280
Dichotomous tree structure 266–269
Direction 4, 175–176, 287
 of becoming 186
 Depth 147
 Simultaneity of 187
Discontinuity 125
Disorder (structural)
Displacement 71, 80, 88–90
 Direction of 90
Disparity 46, 48, 54, 57, 68
Distance 4, 138
 Regular hyper surface 317
Dynamics 71–72, 75, 93, 235, 245

* I am grateful to Francesca De Carli for her assistance with the indexes.

E

Edge detection 126–128, 198
Eikonal 316
Erlangen program (Felix Klein) 318
Epistemology 66
Epoché 169
Evolution 332
Extendedness 201
External world 35, 37, 52, 67

F

Feature
 Distinctive 178
 Distinguishing 178
 Non accidental 334
 Pictorial 140
 Verbal 239
Flash-lag effect 5–7
Flow 115
 Optical 326, 342
Focus 37
 Verbal 236
 Visual 236
Fold 127
Folk psychology 67
Form 304
Forward optics 40
Frame
 Field 119
 Reference 111
 Bundles 324, 329, 342
Frölich effect 5–7, 91, 92

G

Gauge figure 341
Generalised cylinder 330, 343
Genetic art 335
Geometries 318
 Affine 319
 Differential 118–119, 121
 Euclidean 149, 319
 Fractal 271, 321
 Hyperbolic 319
 Isotropic 148–149
 Non-metrical 319
 of Visual flows 129
 of Pictorial space 141, 147
 Ordinal 320
 Projective 149, 320
 Pseudo-metrical 320
 Riemannian 319
 Topological 321

Geons 334

Gestaltpsychologie 7, 167

Gestalt laws

 Closure 248, 262
 Common fate 157
 as grouping 160
 Good continuation
 115–116, 130, 262,
 264–265
 Grouping 155, 157, 282,
 286, 303
 by abstract features 159
 Factor 260, 297
 in abstract domains 161
 Mental 248
 of remembered objects
 157
 Pragnänz 73, 175
 Proximity 248, 263–265
 Similarity 248, 262–265,
 282
 Symmetry 248, 262
Gestaltung 166
 of Direction 174
Graphic equivalents 185, 195

H

Hypercolumn 340
Homogeneous 146
Horizontal transform
 sequence 322
Hypersurface distance 316,
 341

I

Ikebana (lit. living flowers)
 287–289
Illumination 41, 44, 46, 67
Illusion
 Hering 25
 Top hat 156
Image 54–55
 Irradiance equation 325
Imagery 251, 253
Indicatrix 341
Inference 42–43
 Bias 48
 Model 51
Information 116, 316
 Content 36
Inner determinant 180–181
Intentional 21
 Reference 166–167

Intentionality

 Reverse 184, 189
Interrogation in CAD-CAM
 308, 313
Inverse mapping 40
Irradiance 67

J

Japanese
 Garden 259–287
 Painting 274–275

K

Karesansui (lit. dry mountain
 stream) 258, 261–262, 269,
 274, 276–277, 286, 297

L

Landscape design 259
Layer 18–19, 21, 120, 124, 189
 Background 24
 Foreground 24
Level 8, 18–19, 24–26, 161, 166,
 338
 Dependence of 17, 27
 Ontological 189
 Phenomenal 27
Lexical unit 290
Lie group 313
Lightness 37, 45, 52–54, 67
Likelihood 43, 48
 Maximum estimator 44
Lipschitz continuos 317, 342
Long-range connections 119
L-systems 334
Luminance 42, 67, 68

M

Ma (spatiotemporal intervals)
 262
Mathematical morphology
 310, 322
Measurement 51, 53, 66–67,
 138, 257
Medial axis 266–272
Medial scaffold 330
Medial symmetry graphs 327,
 330
Molecular graphs 330
Momentum
 Implied 76
 Representational 5, 76–78

Morphogens 313
 Morphometrics 321
 Morse theory 331, 343
 Motion parallax 37, 46–49,
 54, 57, 68
 Motion field 326
 Movement
 γ 174–175
 Eye 252
 In pictorial space 146
 Of inner parts 174

N

Nihonga (Japanese painting)
 278

O

Object 4, 14, 20–23, 25, 36, 304
 and content 15, 23
 and surfaces 40
 Mental 155
 Natural 20
 Outline 305
 Phenomenal 14, 170
 Identity of 182
 Physical 14, 185
 Pictorial 20, 135
 Presented 16
 Real 99
 Remembered 157
 Stereokinetic 99, 175–176
 Visual 174, 185
 Occlusion 278
 Onset Repulsion Effect 91–92
 Ontology 50
 Optimisations 311
 Optics
 Forward 40
 Geometrical 315
 Inverse 40, 47
 Orientation 36, 115, 119, 172,
 175
 Contour 106
 Osculating circle 118, 323
 Outline 304

P

Parallel transport 323
 Part 21, 22, 23, 176
 Functional 181
 Independent 22, 185
 Inner 175

Non independent 22, 185
 Object 109, 110
 Visual 189
 Visual-verbal 239
 Pattern 37, 78, 351–356, 363
 Reflectance 37
 Perception 35
 as Inference 35, 40, 43,
 46–49, 52
 and Memory 88
 Direct 221
 Ecological 48
 Perceptual
 Conflict 40
 Ground 56, 64
 Identity 62
 Perspective 270–272
 Phase-change interface 35, 36
 Phenomenology 38, 55, 66
 Pictorial relief 140–142
 Picture 137–138, 142
 as psychophysical stimulus
 142
 Description 236, 242, 244
 Reality in 222
 Surface 55
 Viewing 236
 Point of indeterminacy 24
 Presentation 16, 38, 39, 69,
 167, 170, 183
 Abstract 15
 Concrete 15
 Object of 167
 Mental 168, 188
 Perceptual 52–56, 58, 66,
 168, 188
 Primitive 26, 198
 Picture 201, 207–208
 Volumetric 198
 Property
 Geometric 36
 Intentional 297
 Secondary 49

Q

Quality
 Depictive 279
 Expressive 180
 Foundational 165
 Global 180
 Metric 165
 Primary 165, 177
 Secondary 177

Structural 259
 Tertiary 177
 Visual 165, 177

R

Reaction-diffusion 332
 Realism
 Direct 67
 Naïve 52
 Relation
 Building above 1, 19, 166
 Overforming 18–19
 Part/Whole 21, 185, 352,
 363
 Temporal-semantic 237,
 241
 Relaxation labeling 123–124,
 126–128
 Representation 14, 17, 37, 39,
 106, 117, 207
 and design 56
 and gravity 81
 Conflict 39, 54–56
 Field 122
 Neural 111
 of parts 112
 Pictorial 186

S

Sakuteiki (Classical Japanese
 gardening text) 260, 288
 Scaffold
 Medial 330
 Shock 330
 Self-similarity 262–263
 Shading 36
 Shading flow 127
 Shading flow boundaries
 116
Shakkei (borrowed scenery)
 278
 Shape 303–305
 Arboreal 279–287
 as structural field 305
 as form 304
 as outline 303
 Bounding 36
 Dynamic 335
 From shading 325, 331
 Grammar 333
 Codon 338
 Process 335, 337
 Modifier 203

Operator 307
Pictorial 149
Probing 306, 308
Repetition 286
Representation 107, 306
Shock 340
 Graph 328
 Scaffold 330
Transform 307
Singularity 124
 Theory 340
Slant 42, 67
Softness 284
Space 304
 Action 3
 Homogeneous 3
 Illusory 105
 Optical 3
 Neural 105
 Perceptual 7
 Personal 3
 Phenomenal 3, 6, 165–166,
 176
 Pictorial 3, 37, 135, 139,
 146, 165
 Potential 170
 Relief 140
 Structure of 353
 Visual 3
Stereokinesis 66
Stimulus 53, 181, 183
 Frozen action 78
 Inducing 76
Structural skeleton 73,
 171–172, 176, 187
Structural field 305, 307
Structural order
 in literary works 289–290
 in musical composition
 289
Structure

 Anticipatory 182
 Dynamic 179
 Inner causal 184
 Visual 296
Surface 35
 and Design 64
 Articulation 58–59, 61–62
 Continuity 37
 Cues 43
 Discontinuous 127
 Geometry 41–42, 46
 Identity 51–53, 55–56, 64
 Illumination 41, 44
 Lightness 37, 51
 Luminance 41
 Microtexture 37
 Reflectance 37, 41
 r-surface 38–39, 67
 p-surface 38–39, 51–52, 67
 Shape 41
 Stereokinetic 99–103
 Transparent 101
System
 Denotation 195, 198, 201,
 203, 208, 211, 213
 Drawing 195, 198, 208, 211
 Mark 195, 199, 208
 Representational 195, 199,
 216

T

T'ai-chi tu 81–90
Tension 74, 171, 257
Textural laciness 284
Texture 36, 125, 277
Tilt 42, 67
Time of presentness 9, 168,
 181
Transphenomenal correlates
 169

Transform

 Differential 313
 Idempotent 309
 Invertible 308
 Linear 311
 Lorentz 312
 Orthogonal 312
 Separable 311
 Shock 317
 Symplectic 312
 Vertical t. sequences 327
T-junction 129, 208, 212–213
Tree
 Pruning 284–285
 Shape 285–287

V

Vergence 47, 48
Veridicality 25, 43, 68, 136,
 137, 141, 143, 170
Visual
 Balance 261
 Cortex 105, 117
 Primary 117
 Dimensions 106–107
 Direction 150
 Flow 115, 123
 Focusing 236
 Interest 60
 Preference 37
 Reconstruction 36
Voronoi diagram 327

W

Wavefront propagation 315

Z

Zipf's law 288–295

Advances in Consciousness Research

A complete list of titles in this series can be found on the publishers' website, www.benjamins.com

- 68 **BRÅTEN, Stein (ed.):** On Being Moved. From mirror neurons to empathy. x, 313 pp. + index. *Expected March 2007*
- 67 **ALBERTAZZI, Liliana (ed.):** Visual Thought. The depictive space of perception. 2006. xii, 380 pp.
- 66 **VECCHI, Tomaso and Gabriella BOTTINI (eds.):** Imagery and Spatial Cognition. Methods, models and cognitive assessment. 2006. xiv, 436 pp.
- 65 **SHAUMYAN, Sebastian:** Signs, Mind, and Reality. A theory of language as the folk model of the world. 2006. xxvii, 315 pp.
- 64 **HURLBURT, Russell T. and Christopher L. HEAVEY:** Exploring Inner Experience. The descriptive experience sampling method. 2006. xii, 276 pp.
- 63 **BARTSCH, Renate:** Memory and Understanding. Concept formation in Proust's *A la recherche du temps perdu*. 2005. x, 160 pp.
- 62 **DE PREESTER, Helena and Veroniek KNOCKAERT (eds.):** Body Image and Body Schema. Interdisciplinary perspectives on the body. 2005. x, 346 pp.
- 61 **ELLIS, Ralph D.:** Curious Emotions. Roots of consciousness and personality in motivated action. 2005. viii, 240 pp.
- 60 **DIETRICH, Eric and Valerie Gray HARDCASTLE:** Sisyphus's Boulder. Consciousness and the limits of the knowable. 2005. xii, 136 pp.
- 59 **ZAHAVI, Dan, Thor GRÜNBAUM and Josef PARNAS (eds.):** The Structure and Development of Self-Consciousness. Interdisciplinary perspectives. 2004. xiv, 162 pp.
- 58 **GLOBUS, Gordon G., Karl H. PRIBRAM and Giuseppe VITIELLO (eds.):** Brain and Being. At the boundary between science, philosophy, language and arts. 2004. xii, 350 pp.
- 57 **WILDGEN, Wolfgang:** The Evolution of Human Language. Scenarios, principles, and cultural dynamics. 2004. xii, 240 pp.
- 56 **GENNARO, Rocco J. (ed.):** Higher-Order Theories of Consciousness. An Anthology. 2004. xii, 371 pp.
- 55 **PERUZZI, Alberto (ed.):** Mind and Causality. 2004. xiv, 235 pp.
- 54 **BEAUREGARD, Mario (ed.):** Consciousness, Emotional Self-Regulation and the Brain. 2004. xii, 294 pp.
- 53 **HATWELL, Yvette, Arlette STRERI and Edouard GENTAZ (eds.):** Touching for Knowing. Cognitive psychology of haptic manual perception. 2003. x, 322 pp.
- 52 **NORTHOFF, Georg:** Philosophy of the Brain. The brain problem. 2004. x, 433 pp.
- 51 **DROEGE, Paula:** Caging the Beast. A theory of sensory consciousness. 2003. x, 183 pp.
- 50 **GLOBUS, Gordon G.:** Quantum Closures and Disclosures. Thinking-together postphenomenology and quantum brain dynamics. 2003. xxii, 200 pp.
- 49 **OSAKA, Naoyuki (ed.):** Neural Basis of Consciousness. 2003. viii, 227 pp.
- 48 **JIMÉNEZ, Luis (ed.):** Attention and Implicit Learning. 2003. x, 385 pp.
- 47 **COOK, Norman D.:** Tone of Voice and Mind. The connections between intonation, emotion, cognition and consciousness. 2002. x, 293 pp.
- 46 **MATEAS, Michael and Phoebe SENGERS (eds.):** Narrative Intelligence. 2003. viii, 342 pp.
- 45 **DOKIC, Jérôme and Joëlle PROUST (eds.):** Simulation and Knowledge of Action. 2002. xxii, 271 pp.
- 44 **MOORE, Simon C. and Mike OAKSFORD (eds.):** Emotional Cognition. From brain to behaviour. 2002. vi, 350 pp.
- 43 **DEPRAZ, Nathalie, Francisco J. VARELA and Pierre VERMERSCH:** On Becoming Aware. A pragmatics of experiencing. 2003. viii, 283 pp.
- 42 **STAMENOV, Maxim I. and Vittorio GALLESE (eds.):** Mirror Neurons and the Evolution of Brain and Language. 2002. viii, 392 pp.
- 41 **ALBERTAZZI, Liliana (ed.):** Unfolding Perceptual Continua. 2002. vi, 296 pp.
- 40 **MANDLER, George:** Consciousness Recovered. Psychological functions and origins of conscious thought. 2002. xii, 142 pp.
- 39 **BARTSCH, Renate:** Consciousness Emerging. The dynamics of perception, imagination, action, memory, thought, and language. 2002. x, 258 pp.
- 38 **SALZARULO, Piero and Gianluca FICCA (eds.):** Awakening and Sleep-Wake Cycle Across Development. 2002. vi, 283 pp.
- 37 **PYLKKÄNEN, Paavo and Tere VADÉN (eds.):** Dimensions of Conscious Experience. 2001. xiv, 209 pp.

- 36 **PERRY, Elaine, Heather ASHTON and Allan H. YOUNG (eds.):** *Neurochemistry of Consciousness. Neurotransmitters in mind. With a foreword by Susan Greenfield.* 2002. xii, 344 pp.
- 35 **Mc KEVITT, Paul, Seán Ó NUALLÁIN and Conn MULVIHILL (eds.):** *Language, Vision and Music. Selected papers from the 8th International Workshop on the Cognitive Science of Natural Language Processing, Galway, 1999.* 2002. xii, 433 pp.
- 34 **FETZER, James H. (ed.):** *Consciousness Evolving.* 2002. xx, 253 pp.
- 33 **YASUE, Kunio, Mari JIBU and Tarcisio DELLA SENTA (eds.):** *No Matter, Never Mind. Proceedings of Toward a Science of Consciousness: Fundamental approaches, Tokyo 1999.* 2002. xvi, 391 pp.
- 32 **VITIELLO, Giuseppe:** *My Double Unveiled. The dissipative quantum model of brain.* 2001. xvi, 163 pp.
- 31 **RAKOVER, Sam S. and Baruch CAHLON:** *Face Recognition. Cognitive and computational processes.* 2001. x, 306 pp.
- 30 **BROOK, Andrew and Richard C. DEVIDI (eds.):** *Self-Reference and Self-Awareness.* 2001. viii, 277 pp.
- 29 **VAN LOOCKE, Philip (ed.):** *The Physical Nature of Consciousness.* 2001. viii, 321 pp.
- 28 **ZACHAR, Peter:** *Psychological Concepts and Biological Psychiatry. A philosophical analysis.* 2000. xx, 342 pp.
- 27 **GILLET, Grant R. and John McMILLAN:** *Consciousness and Intentionality.* 2001. x, 265 pp.
- 26 **Ó NUALLÁIN, Seán (ed.):** *Spatial Cognition. Foundations and applications.* 2000. xvi, 366 pp.
- 25 **BACHMANN, Talis:** *Microgenetic Approach to the Conscious Mind.* 2000. xiv, 300 pp.
- 24 **ROVEE-COLLIER, Carolyn, Harlene HAYNE and Michael COLOMBO:** *The Development of Implicit and Explicit Memory.* 2000. x, 324 pp.
- 23 **ZAHAVI, Dan (ed.):** *Exploring the Self. Philosophical and psychopathological perspectives on self-experience.* 2000. viii, 301 pp.
- 22 **ROSSETTI, Yves and Antti REVONSUO (eds.):** *Beyond Dissociation. Interaction between dissociated implicit and explicit processing.* 2000. x, 372 pp.
- 21 **HUTTO, Daniel D.:** *Beyond Physicalism.* 2000. xvi, 306 pp.
- 20 **KUNZENDORF, Robert G. and Benjamin WALLACE (eds.):** *Individual Differences in Conscious Experience.* 2000. xii, 412 pp.
- 19 **DAUTENHAHN, Kerstin (ed.):** *Human Cognition and Social Agent Technology.* 2000. xxiv, 448 pp.
- 18 **PALMER, Gary B. and Debra J. OCCHI (eds.):** *Languages of Sentiment. Cultural constructions of emotional substrates.* 1999. vi, 272 pp.
- 17 **HUTTO, Daniel D.:** *The Presence of Mind.* 1999. xiv, 252 pp.
- 16 **ELLIS, Ralph D. and Natika NEWTON (eds.):** *The Caldron of Consciousness. Motivation, affect and self-organization — An anthology.* 2000. xxii, 276 pp.
- 15 **CHALLIS, Bradford H. and Boris M. VELICHKOVSKY (eds.):** *Stratification in Cognition and Consciousness.* 1999. viii, 293 pp.
- 14 **SHEETS-JOHNSTONE, Maxine:** *The Primacy of Movement.* 1999. xxxiv, 583 pp.
- 13 **VELMANS, Max (ed.):** *Investigating Phenomenal Consciousness. New methodologies and maps.* 2000. xii, 381 pp.
- 12 **STAMENOV, Maxim I. (ed.):** *Language Structure, Discourse and the Access to Consciousness.* 1997. xii, 364 pp.
- 11 **PYLKKÖ, Pauli:** *The Aconceptual Mind. Heideggerian themes in holistic naturalism.* 1998. xxvii, 297 pp.
- 10 **NEWTON, Natika:** *Foundations of Understanding.* 1996. x, 211 pp.
- 9 **Ó NUALLÁIN, Seán, Paul Mc KEVITT and Eoghan Mac AOGÁIN (eds.):** *Two Sciences of Mind. Readings in cognitive science and consciousness.* 1997. xii, 490 pp.
- 8 **GROSSENBACHER, Peter G. (ed.):** *Finding Consciousness in the Brain. A neurocognitive approach.* 2001. xvi, 326 pp.
- 7 **MAC CORMAC, Earl and Maxim I. STAMENOV (eds.):** *Fractals of Brain, Fractals of Mind. In search of a symmetry bond.* 1996. x, 359 pp.
- 6 **GENNARO, Rocco J.:** *Consciousness and Self-Consciousness. A defense of the higher-order thought theory of consciousness.* 1996. x, 220 pp.
- 5 **STUBENBERG, Leopold:** *Consciousness and Qualia.* 1998. x, 368 pp.
- 4 **HARDCASTLE, Valerie Gray:** *Locating Consciousness.* 1995. xviii, 266 pp.
- 3 **JIBU, Mari and Kunio YASUE:** *Quantum Brain Dynamics and Consciousness. An introduction.* 1995. xvi, 244 pp.

- 2 **ELLIS, Ralph D.:** Questioning Consciousness. The interplay of imagery, cognition, and emotion in the human brain. 1995. viii, 262 pp.
- 1 **GLOBUS, Gordon G.:** The Postmodern Brain. 1995. xii, 188 pp.