

in: H. Hecht, R. Schwartz & M. Atherton (eds.) (2003). *Looking into Pictures: An Interdisciplinary Approach to Pictorial Space* (pp. 17-60). Cambridge, Mass.: MIT Press

Conjoint Representations and the Mental Capacity for Multiple Simultaneous Perspectives

Rainer Mausfeld

Table of Contents

- The dual character of pictures
- Conflicts in cue integration with respect to depth or spatial representations
- Conjoint representations in picture perception
- Triggering and parameter setting: The dual function of sensory inputs with respect to representational primitives
- Conjoint representations as a general structural property of our basic cognitive architecture
 - Examples from visual psychophysics
 - Examples from other areas of cognitive science
- Vagueness, smooth transitions between representational primitives, and the need of a 'proximal mode'
- References

Mens videt, mens audit: Cetera surda et coeca.
It is the mind that sees and the mind that hears;
the rest are deaf and blind.
*Epicharmos*¹

Common-sense taxonomies were, inevitably, the origin from which the natural sciences, at their earliest stages of development, derived their categorizations of phenomena. This can be witnessed by the classical division of physics into e.g. optics, acoustics, theory of heat, and mechanics. During the process of its theoretical development, physics became increasingly divorced from these kinds of classifications and rather grouped phenomena in accordance with its own internal theoretical structure (the classical theory of heat, for instance, disintegrated into statistical mechanics, on the one hand, and electrodynamics, on the other hand). In perceptual psychology corresponding pre-theoretical classifications of phenomena are mirrored in the standard textbook organization in terms of salient perceptual attributes, such as colour, depth, size, or form. In this field, it will likely prove to be more difficult than it was in physics to dispense with common-sense classifications of phenomena and to instead follow lines of theorizing that are traced out by the development of successful explanatory accounts. This difficulty is due to the power that the phenomenal appearance exercises over our way of theoretically grouping perceptual phenomena. Although we are well aware that common-sense taxonomies are an inapt guide for the endeavour to achieve, within the framework of the natural sciences, a theoretical understanding of the mind, we are held captive by the appearances. We are inclined to believe that perception works in the way it phenomenally appears to us and that a theoretically fruitful classification of phenomena basically follows our common-sense psychological intuitions. It were such pre-theoretical common-sense taxonomies that in perceptual psychology brought forth sub-fields such as picture perception or colour perception.

¹ Diels (1922, fr. 12, p. 123)

For someone, like me, who is working in the field of colour perception, picture perception does not appear as a natural field of inquiry, all the more so because hardly any other two fields in visual perception are as remote in their approaches and their theoretical frameworks as picture perception and colour perception. Nevertheless, I believe that phenomena in these fields share, on a more abstract level, structural similarities that appear to point to deeper principles of our mental architecture. This is hardly surprising, because there is no reason to expect that classes of perceptual phenomena that arise in the context of certain artefacts, such as pictures, and classes of phenomena that pertain to certain phenomenal attributes, such as colour, will survive as theoretically useful categories when more successful explanatory accounts of perception have eventually been developed. Rather, such accounts would result in a classification of phenomena that is determined by the actual internal principles underlying perception, whatever these may turn out to be.

The most promising general approach to perception appears to me to be one that follows ethological lines of thinking. Such approaches have, notably when couched in computational terms, already yielded intriguing explanatory frameworks of promising range and depth. From an ethological perspective, the core task of perceptual psychology is to investigate the structure of perceptual representations and the nature of the representational primitives on which it rests. With respect to colour perception such an approach has brought forth what I believe to be interesting theoretical speculations about the nature of the representational primitives and about the internal structure they give rise to. The internal principles that are suggested by these speculations and which we are only beginning to understand are fairly general ones that cross-cut common-sense categorizations of phenomena. Although not much is known presently about the structure of the representational primitives to which these principles refer, some of their general properties are suggested by interesting structural similarities that certain phenomena from various domains in cognitive science appear to share. Among these phenomena is our ability to perceptually deal with the so-called dual character of pictures. This ability is, I believe, part of a general structural property of our basic cognitive architecture that refers to the handling of *conjoint representations* over the same input. In various domains and at different levels of the cognitive system we can encounter phenomena that are likely due to the internal handling of multiple conjoint and often competing representations. From this point of view, picture perception and its dual character are a special instance where we exploit these given capacities in the context of human artifacts. Before arguing for this view in greater detail, I will briefly delimit the topic of my inquiry.

In picture perception, more than, say, in research on stereo vision or colour coding, the tension looms large between what are to be considered universal and what cultural and conventional aspects, a tension that mirrors and maintains the time-honoured distinctions that placed physis and ethos, nature and convention, essential and accidental properties in well-nigh irreconcilable opposition to one another. The corresponding issues are a matter of much debate between - to use some fashionable jargon - universalists and cultural relativists. Outside some areas of vision and language, little of substance is presently known about where to draw a dividing line between universal aspects and aspects of cultural variation, individual plasticity or learning history. However, the entire idea of cognitive science rests on the assumption that *some such* distinction can be drawn at all. This also holds for inquiries centring upon picture perception. The perception of pictures, on the basis of multiple pictorial components of very different status, involves highly complex interactions of our perceptual faculty and various interpretative faculties, which are presently not understood very well. These interactions give rise to a high degree of cultural variation. I shall deliberately ignore the cultural dimension of picture perception and, with respect to the so-called dual character of pictures, focus on structural elements of perception that seem to be part of our basic cognitive endowment.

The Dual Character of Pictures

Pictures and pictorial representations, though highly impoverished two-dimensional abstractions of what is depicted, can evoke strong perceptual impressions of objects, spatial relations or events within us. A dominant theme in the field of picture perception has been the issues centring upon notions of perceptual space and the extent to which corresponding percepts can be evoked by features of pictorial representations, notably through linear perspective (cf. Haber, 1980; Rogers, 1995). During the Renaissance an increasing interest emerged in techniques of linear perspective. This was motivated by the artists' desires to imitate nature and to achieve 'visual truth' in their paintings. This idea gave rise to related inquiries into artistic techniques for the evocation of space and, in particular, into techniques for creating geometrically correct two-dimensional pictorial representations on a canvas of the three-dimensional layout of the pictured scene. In such investigations, as Kemp (1990, p.165) observed, "the eye figures little, the mind features even less." Rather what was to be accomplished was "the dem-

onstration of an internally consistent system of the spatial elements in a picture and, above all, a proof that the system rested upon non-arbitrary foundations." (ibid., p. 11) The canvas was regarded as a window, often referred to today as an *Alberti window*, through which the painter views the world and which intersects his visual cone (Lindberg, 1976). This gave rise to the idea that a realistic appearance of depth and space can be achieved in pictures by mimicking the exact geometrical relations in the structure of light that reaches the eye from a three-dimensional scene. Consequently, a system of construction rules, in the sense of artistic engineering techniques for the purpose of creating, on a flat canvas, pictorial representations that induce a strong appearance of depth in the observer, gained prominence in Renaissance art.² Although these artistic techniques later joined with ideas on geometrical processes of image formation in the eye, their use and development were primarily shaped by considerations internal to the complex variety of cultural purposes underlying artistic productions. However, for the endeavour to imitate nature and to achieve visual truth in two-dimensional representations of the world, the importance of rules for linear perspective is on a par with those for simulating the effect of lights and the interaction of light and objects by using spatial pigment patterns on a flat surface (Schöne, 1954). It is a historically contingent development of art history that resulted in linear perspective, rather than other aspects, first gaining prominence in this context.

The notion of a *dual character of pictures* basically refers to the phenomenon that pictures can generate an in-depth spatial impression of the scene depicted while at the same time appearing as flat two-dimensional surfaces hanging on a wall. Michotte (1948/1991) recognized the challenge that this kind of phenomenon poses for perception theory.³ A description in terms of a perceptual conflict between the perceived flatness of the picture's surface and the perceived depth of what is depicted captures only a small fraction of the perceptual enigmas involved. By careful phenomenological obser-

² Also pictorial devices - flatening techniques (Willats, 1997) - have been developed to control the perceptual balance between the flatness aspect of the picture and the depth aspect of what is depicted, such as accidental alignments between two or more parts of a scene and the position of the viewer, the use of mixed and mutually inconsistent perspectives, or obtrusive surface marks (cf. also Michotte, 1962, p. 515).

³ These phenomena could, from a physicalistic perspective, be described as a kind of discrepancy between what is physically there, viz. a flat surface, and the perceptual impression evoked. However, framing the problem this way amounts to conflating the level of the physical generation process of the sensory input with the level referring to perceptual mechanisms by which this sensory input is exploited (cf. Mausfeld, 2002).

vations of viewing pictures one can easily gather indications of how difficult it is to precisely describe at all what the percepts and thus what the perceptual achievements are in these situations.⁴ I will deliberately leave out many of the problems that we encounter in describing what exactly the percepts or the perceptual achievements in corresponding situations of viewing pictures are. I will rather distinguish only two general problems to which the notion *dual character of picture* seems to refer, which I will address in turn. The first is the problem of cue integration with respect to depth or, more generally, spatial representations.⁵ The second is a problem that seems to me to be even more complex and much deeper, namely the problem of what I refer to as *conjoint representations over the same inputs* and our ability to smoothly handle them. By conjoint representations I will, in an intuitive and tentative way, denote two or more representational primitives of the same type or of different types that exploit the same input properties in different and interdependent ways. The special case of *competing* conjoint representations is furthermore characterized by the property that the parameters of one representational primitive that relate to a certain aspect of the input are antagonistically coupled with parameters of the other representational primitive that refer to the same aspect. As an illustration think of a surface viewed under some chromatic illumination: the incoming colour signal is internally exploited in terms of two components, viz. 'surface colour' and 'illumination colour' (to be understood as internal, and not as physical concepts); the extent to which a 'surface' representations exploits the incoming colour

⁴ Phenomenological observations that appear particularly salient or enigmatic do not necessarily have a particular relevance for perception theory. Although phenomenological observations of various kinds are of prominent heuristic importance for perception theory, they do not carry a kind of 'epistemological superiority'. Phenomenological observations do not provide a 'direct access' to the nature of representational primitive; they rather result from an interplay of various faculties, including linguistic and interpretative ones. Thus they are, within a naturalistic inquiry into the principles of perception, on a par with many other sources that provide relevant facts and observations.

⁵ The notion of 'representation' is burdened with a high degree of ambiguity due to its multifarious meanings. Many corresponding locutions in this paper, such as 'pictorial representation', refer to ordinary discourse. With respect to these, I do not attach much importance, in the present context, to carefully distinguishing different meanings. In the context of explanatory frameworks for certain phenomena, however, I use the notion of 'representation', e.g. in the terms 'spatial representations' or 'surface' representations, to denote elements of postulated internal structure that are part of an inference to the best explanation. In the context of perception theory, this neither involves some particular ontological commitments nor any reference to the external world. More concretely, a 'surface' representation is not, in any meaningful sense, to be understood as a representation *of* physical surfaces. Dispensing with notions of 'reference', 'truth' or 'veridicality' within explanatory frameworks of perception theory is, as, with respect to naturalistic inquiries of the mind, Chomsky (1996; 2000) has argued most convincingly and adamantly, entirely in line with standard methodological principles of the natural sciences.

signals in terms of its colour parameter antagonistically constraints the values that an 'ambient illumination' representation can assign to its colour parameter, and vice versa.

The two issues of the integration of conflicting cues with respect to a specific representational primitive, on the one hand, and a dual representation that results from conjoint representations over the same inputs, on the other hand, are often conflated.

Conflicts in Cue Integration with Respect to Depth or Spatial Representations

The lines along which we can theoretically explore the dual character of pictures considered as a case of cue integration are comparatively well explored in visual psychophysics. Although problems of cue integration are intricate enough we have both a wealth of experimental results and subtle mathematical modelling tools, such as the Bayesian framework, which has proved a fruitful basis for dealing with these issues. All this is well-known and I will only briefly review some of the relevant evidence in order to be able to better differentiate the second problem of conjoint representations from issues of cue integration.

We know from psychophysics that the depth one experiences, i.e. the apparent variation in surface relief and the relative location of objects in 3D-space, is constructed from multiple sources of evidence. These cues can carry different weight with respect to internal spatial representations, weights that can be, but are not necessarily in line with natural ecological constraints. For instance, occlusion carries a strong internal weight, but provides only weak constraints for spatial representations, namely that the depth on one side of the border is greater than on the other side. Another example is shading, which provides information about the surface normal at each location. Even cues that are unrelated to depth per se can be used to disambiguate other cues, such as spatial frequency content for 3D-shape-from-texture cues. If different sources of evidence favour different and mutually incoherent spatial representations of the same input, the visual system often has some kind of default preference for solving this cue conflict. It does this without providing any phenomenal access to alternative representations. A case in point is the Ames room where systematic manipulations in linear and texture perspective cues result in estimates of depth and size which are based on a cue integra-

tion that vetoes or ignores cues of familiar size; the cue integration seems rather to be in line with internal heuristics such as '*lines which are nearly parallel in the image are parallel in 3D*', or '*lines which meet at a common vertex in the image also meet at a common vertex in 3D*'.

In picture perception we find a plenitude of ways for the evocation of space and depth on a two-dimensional surface that mirror the plenitude of depth cues. We encounter a similar situation for the integration of *conflicting* cues, namely the integration of stereo disparity information and various monocular depth cues.⁶ While the depth information of the picture surface is perfectly in line with the depth information from the frame or the surrounding wall, the scene depicted can invoke the impression of phenomenally extending in depth.

In principle we could conceive a visual system which integrates these cues in such a way that the stereo information completely dominates and, in the case of incoherence, vetoes monocular information, which phenomenally would result in a reduction of the - already not very strong - 3D-vividness of the scene depicted to zero, i.e. a flat picture without any indication of surface orientation or relative 3D- locations of objects. Interestingly enough, our visual system uses almost the opposite strategy: it is a well-known result from visual psychophysics that monocular cues often dominate the resulting 3D-interpretation over stereopsis, even at close range where stereopsis is most accurate. I will mention only a few corresponding studies. For instance, in a pioneering study, Schriever (1925) found that perspective alone as well as occlusion could overrule disparity information. Schriever also made a wealth of careful observations about attentional effects, vagueness and instability as well as individual differences. Turhan (1937) found that in centre-surround situations brightness gradients of opposite direction can result in perceptual impressions that violate the physical depth relations of infield and surround as provided by disparity information and motion parallax (infields whose brightness gradient has the same direction as that of the surround appear to lie in the same depth level as the surround, while they appear to lie in a different depth plane and often look bent in the case of opposite brightness gradients). Turhan observed that

⁶ Demonstrations for different kinds of situations of conflicting depth cues have been provided, for example, by Hornbostel (1922), in his striking demonstration that the image of a rotating wire cube viewed in a mirror undergoes a dynamic non-rigid transformation (cf. also Adams & Haire, 1959), by Ames (Ittelson, 1952) or by Epstein (1968), Gillam (1968), Youngs (1976), Rogers & Collett (1989), Trueswell & Hayhoe (1993), or Koenderink, van Doorn & Kappers (1994).

physically incompatible depth interpretations of the infield can occur at the same time and often are accompanied by some kind of vagueness of the perceptual impression.

Yellott (1981) showed that an inside-out face - a mould of a face - looks right side out as long as shading is present despite the presence of contradictory disparity information as provided by a random-dot stereogram projected onto the surface of the mask. If the mould is presented solely as a random-dot stereogram with no shading, it is seen inside out, i.e. consonant with the disparity information. Another interesting example was provided by Prazdny (1986), who described a random-dot stereo cinematogram that portrays a flat object in front of a background which changes its two-dimensional shape consistent with a three dimensional rotating wire object, while the binocular disparities were incompatible with the relative depth information specified by the image motions. Due to the appearance of three-dimensionality in these displays he concluded that the kinetic depth effect effectively vetoes the stereo disparity cue with respect to the shape of the object (however, disparity determined the position of the object with respect to the background).

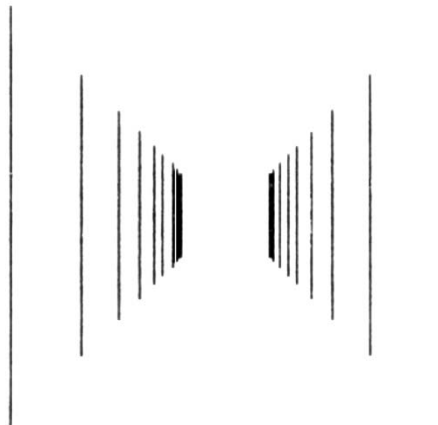


Figure 1: Stimulus configuration used by Stevens and Brookes (1988). The lines are stereoscopically presented as being coplanar, i.e. they increase linearly in disparity from left to right. The 3D impression, however, is of a corridor extending in depth, as suggested monocularly.

A particularly effective demonstration of the monocular influence over stereo information was provided by Stevens and Brookes (1988; see also Stevens, Lees and Brookes, 1991). The lines in the stereoscopically presented figure 1 are coplanar, that is they increase linearly in disparity from left to right. The 3D-impression, however, is of a corri-

dor extending in depth, bordered on either side by columns of vertical lines or stakes. In the stereo-apparatus the innermost lines on either side of the vertical meridian had stereo disparities of ± 11 minutes of arc; the outermost lines had disparities of $\pm 51'$. It is remarkable that the line with $- 11'$ disparity appeared more distant than the line of disparity $+ 51'$. Stevens and Brookes found empirical evidence that stereopsis extracts 3D surface information only where the second spatial derivatives of disparity are non-zero, corresponding to loci where the surface is curved, creased or discontinuous.

We now can directly apply their result to the situation of picture perception. A picture hanging on a wall has no local disparity differences over the picture surface, but only a continuous uniform gradient of disparity indicating a flat although extended surface. Accordingly, because of the specific properties of the human visual system for integrating binocular and monocular sources of depth evidence, as exemplified by Stevens and Brookes' results, stereopsis is particularly weak under these conditions. Our phenomenal spatial impression is therefore - in the presence of monocular depth cues such as occlusion, texture, shading or perspective - determined by these monocular cues. Thus, in picture perception, we take advantage of the internal coding property that depth is derived from disparity only where the surface exhibits continuous curvature or sharp discontinuities, because some binocular disparity information must be discounted in picture perception to interpret drawings at all.

I will not further dwell on the cue integration aspect of the dual character of pictures (for a more recent study of cue integration with respect to depth, see Landy et al., 1995). What I wanted to show is that the conflict between stereo information and monocular depth cues, as well as the corresponding observation that the flatness of the picture plane does not impede the depth impression of the scene depicted, is simply a special case of the specific way our visual system integrates disparity information with other depth cues, as mirrored in a great variety of psychophysical results.

Conjoint Representations in Picture Perception

I will now turn to the second aspect to which the notion of dual character of pictures refers, namely the issue of *conjoint representations over the same input* and our ability to handle them smoothly. This issue is much more puzzling and much less well understood than the conflicting cue aspect. I shall first describe the corresponding kind of phenomenon, as referred to in the literature on picture perception. Like the cue integration aspect, it is, in my view, not specific to picture perception but rather a general and essential property of our cognitive organization.

In picture perception, we can simultaneously have the phenomenal impression of two different types of objects, each of which seems to thrive in its own autonomous spatial framework⁷, namely, on the one hand, the picture surface as an object - with corresponding object properties such as orientation or depth - and, on the other hand, of the depicted objects themselves with their idiosyncratic spatial properties and relations. We seem to have two mutually incompatible spatial representations at the same time; at least in the sense that they are available internally and we can, without any effort, switch to and fro between them.

Before venturing some more speculative ideas about some general properties of our cognitive architecture on which this ability rests, I will list some observations concerning this dual nature of pictures that I consider to be of particular relevance.

In an aside, in order to avoid potential misunderstandings I would like to emphasize that pictorial art in general cannot simply be understood as a kind of frozen optical array or a static boundary case of the optical structure of the input from a scene. Pictorial art is much richer than naturalistic artistic productions and serves a great many different symbolic functions. Pictures are not surrogates for scenes, nor can they be subjected to a criterion of some absolute notion of veridicality (whatever that means) with respect to the scene depicted. Thus, those aspects of picture perception that we potentially can understand from core perceptual principles - and, as I said in the beginning, it is only this part that I will address here - is, from the perspective of cultural studies, the least interesting aspect of picture perception. What we usually refer to when we talk about picture perception are symbolic interpretations at various levels, and thus aspects that

⁷ This is conspicuously illustrated and symbolized by Magritte's painting *La condition humaine* which plays with the many different levels involved.

pertain to highly complex interactions of our perceptual faculty and various interpretative faculties. Within the framework of the cognitive sciences we virtually know next to nothing about the cognitive principles underlying these achievements.

1. A continuous path of transitions exists from a view of a real 3D-scene to a scene as depicted (or abstracted) on a canvas

We can easily construct a continuous path from the view of a real 3D-scene to a real 3D-scene viewed as a kind of frozen Alberti-window to a photo of this frozen optical array and to a highly reduced drawing of the relevant contours of the scene. This allows us to experimentally investigate all sorts of transitions and boundary conditions in picture perception.⁸

2. We can phenomenally accentuate one or the other aspect and switch back and forth in an effortless way

This aspect is phenomenally so conspicuous and striking that we usually do not pay much attention to it. It is at the core of what we mean when we refer to the dual nature of pictures. Though such switches are correlated with depth aspects, they actually pertain to the entire perceptual organization of the visual field and thus to attributes like shape, or shading and brightness gradients. A wealth of observations pertaining to this kind of phenomenon have been reported in the literature, a wealth that oddly contrasts with the silence about what to make of these observations theoretically.

Gombrich (1982) made the important observation that one has to achieve the proper mental attitude to take full advantage of the capacity to switch back and forth between the reality of the picture as an object and the reality of the depicted objects. Because of this, people at earlier stages of cultural development regularly seem to have problems in seeing what is depicted in a photo. For instance, Deregowski, Muldrow and Muldrow (1972) reported that people from a remote Ethiopian tribe when presented with a drawing of an animal would pay attention to the characteristics of the drawing paper

⁸ The most extreme versions are trompe-l'oeil paintings that aim to induce the viewer to perceive the painted object as reality. Such illusionistic effects only work in cases where the painted 3D- perspective is as shallow as possible, which explains the highly restricted pictorial themes of trompe-l'oeil paintings (cf. Mauries, 1997).

but would ignore the picture; they exhibited a complete inattention to the content of the representation while concentrating on the medium. Several others have reported essentially the same observation. When people at earlier stages of cultural development regularly have problems in seeing what is depicted on a photo, they have not yet attained the ability to exercise what Gombrich called the proper mental attitude and thus cannot fully exploit a given cognitive capacity in the case of previously unknown artefacts. However, as Hagen & Jones (1978, p. 192) concluded from a review of corresponding studies, "this coexistence of information poses few problems even for the naive observers when pictures represented only single solid objects. There is no evidence whatsoever that any group of people see pictures of faces, cups, hunters, antelopes or elephants as flat 'slices of life', as it were." Many other interesting regularities were found with respect to the ability to simultaneously handle both types of reality, as it were. For instance, outline drawings present fewer difficulties than photos to naive observers; thus contour information seems to be of greater importance than texture.⁹

Corresponding observations are, of course, not confined to picture perception, but pervade psychophysics and perceptual psychology. For example, in the study mentioned previously, Stevens and Brookes (1988, p. 383) made the observation that experienced stereo observers can also discern the true stereo depth of the component lines with scrutiny, as if they can selectively disregard the monocular depth interpretation.

3. The 'realities' of pictures as objects and depicted objects bear different amounts of internal computational relevance and phenomenological vividness

We can switch back and forth between these two kinds of spatial representations but this is not a switch between a 2D-representation (within some 3D-representation) and a different fully-fledged 3D-representation. Rather, the spatial representation of the scene depicted is phenomenally quite shallow. In a sense we could regard it as a phenomenal analogue to Marr's 2 1/2 D spatial representations, that is, what we perceptually experience are local surface orientation, distance from viewer, or discontinuities in depth and discontinuities in surface orientation, without either the phenomenological vividness or the other internal properties of a full-blown 3D- representation. There are, of course,

⁹ Polanyi (1970) and Pirenne (1970) introduced the terms "focal vs. subsidiary awareness" to deal with these observations: Focal awareness refers to the subject represented, while subsidiary awareness refers to the characteristics of the surface of a picture.

several other aspects that are responsible for the reduced 3D-vividness of the scene depicted, notably a lack of cues provided by motion, and the ranges of colour and luminance contrast, which are much narrower than those of real scenes.

Even if we attend to the two-dimensional picture surface we may experience the depicted object, say, a line drawing of a cube, in a mandatory way as three-dimensional, and yet this 3D spatial representation lacks other crucial elements of a full-blown 3D-representation; for instance it would hardly fool us into trying to grasp and rotate the cube.¹⁰ Both types of representation exhibit therefore quite different kinds of anchoring within the internal computational structure, and we can safely conjecture that we have specific mechanisms subserving 'flat' representations by ignoring certain aspects of 3D-structure. Pieter Saenredam's *Interior of the St Jacob Church in Utrecht* may serve as another illustration. In this case the conflict with size relations in the spatial representation of the picture viewed within the environmental context already suffices to divorce the - otherwise mandatory - 3D-interpretation of the scene depicted from other internal coding properties of fully-fledged 3D- representations.

¹⁰ For similar reasons, Michotte (1948/1991, p. 181) emphasised the distinction between *phenomenal three-dimensionality* and *phenomenal reality*. He argued "that three-dimensionality and reality are different properties of our perceptions and must be considered as *independent dimensions* of our visual experience. ... By *reality* we mean an empirical characteristic, the potential for being manipulated. By *three-dimensionality* we mean another empirical characteristic, the capacity for being matched to the volume of a substantial object."



Figure 2: Pieter Jansz Saenredam, *Interior of the St Jacob Church in Utrecht*, 1642, Alte Pinakothek, München

4. The two representations are not independent but interlocked

When we look at a picture like the one displayed, it seems that we can achieve a kind of autonomous spatial representation of the scene depicted that is detached from the normal spatial representation of our environment, including the picture surface plane. In other words, we seem to use our faculty for spatial representations to emulate certain of its achievements in a restricted local framework.¹¹ As a result of which, the internal out-

¹¹ Even *within* a single picture, different and globally incoherent local spatial representations can be elicited. This was deliberately employed to create certain aesthetic effects, as in Piero della Francesca's painting *The Resurrection of Christ* (see Field, 1993) or paintings by de Chirico.

ternal output of such an emulation is a partly autonomous, though comparatively faint local spatial representation within the canonical spatial representation of our environment. In fact, both representations are nevertheless internally interlocked; just *how* they are interlocked is only poorly understood. I will mention only two observations in this regard: firstly, in many cases, we can experience that - at least locally - the spatial vividness that is gained by one aspect is lost by the other. Second, as Hagen & Jones (1978, p. 194) also pointed out, "the space behind the picture plane is not completely separated from the space of ordinary environment which surrounds the picture." For instance, Deregowski, Muldrow and Muldrow (1972) observed an interesting effect of a horizontal vs. a vertical presentation of a picture of a profiled standing buck. When the picture was presented lying flat on the ground most observers reported that the buck was lying down, when it was presented vertically, they reported that the buck was standing up.

The proposals and corresponding observations by Pirenne (1970), Farber & Rosinski (1978), Kubovy (1986), and others that the surface characteristics of the picture have to be available internally, in order to allow some kind of compensation process that corrects distortions caused by an inappropriate viewing geometry, also indicate that both representations are interlocked in complex and poorly understood ways. In other areas, such as colour or brightness perception, we have a better theoretical understanding about how conjoint representations are interlocked.

The structural perceptual properties that we can identify in theoretical analyses of phenomena centring around the dual character of pictures cannot simply be regarded as kinds of 'perceptual irregularities' that are due to encountering an artefactual situation. But rather these phenomena seem to point, in a particularly conspicuous way, to a general perceptual capacity to deal with conjoint representations.

Triggering and Parameter Setting: The Dual Function of Sensory Inputs with Respect to Representational Primitives

In perception theory we can roughly distinguish three aspects of architecture with respect to the relation of sensory inputs and internal representations. i) Several kinds of input properties are exploited by the same kind of internal representation (e.g. computational theories of cue integration), ii) the same input property is independently ex-

ploited by several representations (e.g. in bees, colour vision proper and wavelength-dependent behaviour coexist and subserve independent representations¹², cf. Goldsmith, 1990), and iii) the same type of input property is exploited by *conjoint representations* over the same input (i.e. the *same* input can give rise to several *different* but interlocked output codes and to multiple simultaneous layers of representations). This last type of architecture can be expected to play a prominent role in highly versatile and complex perceptual systems that have to simultaneously subserve a great variety of tasks. Such systems must internally have the outputs of many sub-systems available for purposes of a great variety of higher-order representations and thus have to provide computational means to handle *conjoint representations* over the same input.

In this section I will briefly sketch a theoretical perspective on the structure of perceptual representations that I believe to be promising, both theoretically and empirically, for attempts to deal with conjoint representations. This approach is inspired by two general perspectives, which are related in some of their core ideas. Firstly, by the ideas underlying classic ethological approaches, notably of v. Uexküll, Lorenz and Tinbergen, and by the extension of these ideas to richer and more complex functions (e.g. Wehner, 1987; Marler, 1999; Gallistel, 1998) than those that were studied in earlier years under the heading of 'innate release mechanism'.¹³ An ethological perspective has, in its basic theoretical assumptions, also gained support from computational approaches. Secondly, by Chomsky's (e.g. 2000) *internalist* inquiries into the nature of language and mind (which adhere to the maxim that in rational inquiries into mental phenomena there is no reason to deviate from the methodological principles routinely employed in other domains of the natural sciences with respect to other 'natural objects'; a maxim that should be uncontroversial, yet, in the cognitive sciences, assumptions to the contrary still remain highly influential). Needless to say, the theoretical picture of the basic principles underlying perception that has been emerging in corresponding studies is still very skeletal and, of necessity, has to be based on considerable theoretical speculation. Yet, an ethology-inspired internalist approach, which focuses attention on the structure of the representational primitives underlying perception, seems to provide a very promising framework for asking novel and potentially fruitful questions about the internal architecture of perception. Such an approach is also less susceptible to the

¹² The action spectra for wavelength-dependent behaviour underlying bees' celestial orientation and navigation, depend on more than one pigment, without exhibiting metameric classes, whereas trichromatic color vision is exclusively employed in feeding and recognition of the hive.

¹³ Interestingly enough, Gombrich (1982) also explicitly based his position on ethological arguments.

physicalistic trap (Mausfeld, 2002) which, in the history of perceptual psychology, has often hindered appropriate questions being asked.

In perceptual psychology, a wealth of empirical and theoretical evidence has been marshalled by Gestalt psychology, Michotte's "experimental phenomenology"¹⁴, ethology, studies of the newborn and young children ethology, and computational approaches that indicates that the structure of internal coding is built up in terms of a rich set of *representational primitives*. According to the theoretical picture that has emerged from corresponding studies, perception cannot be understood as the 'recovery' of physical world structure from sensory structure by input-based computational processes. Rather, the sensory input serves as a kind of sign for biologically relevant aspects of the external world that elicits internal representations on the basis of given representational primitives. (Thus, even 'highly impoverished' sensory inputs can trigger perceptual representations whose 'complexity' far exceeds that of the triggering stimulus and whose relation to the sensory input can be contingent from the point of physics or geometry). Although the sensory input is a causally necessary requirement for perceptual representations, the perceptual computations triggered are under the control of an internal programme based on a set of representational primitives. They are representation-driven rather than stimulus-driven.

With respect to human perceptual capacities, this theoretical perspective and the evidence on which it is based suggest distinguishing, as an idealization, a sensory system from a perceptual system.¹⁵ Whereas the sensory system deals with the transduction of

¹⁴ Michotte was particularly sensitive to the problem of meaning in perceptual theory, which he regarded as being intrinsic to the structure of primitives that underlie perceptual organization and that "prefigure" the phenomenal world. "Our research suggests that this primitive structure occurs in the form of a world of 'things' that are separate from one another and are either passive or else animate bodies ultimately endowed with specific 'vital' movements. ... This possibly throws some light on the origin of these concepts, but we ought to stress the biological importance of such spontaneous organization of the phenomenal world since only such organization could enable the individual (whether human or animal) to adapt its reactions before any individual experience had the opportunity to provide it with any structure." (Michotte, 1954/1991, p.45)

¹⁵ This distinction is different in character from widely made distinctions between so-called earlier or lower-level systems and higher-level systems. The latter basically correspond to the sensation-perception distinction as used by Spencer, James, Wundt or Helmholtz, which refers to an alleged hierarchy of processing stages by which the sensory input is transformed into 'perceptions'. In contrast, the present distinction refers to two categorially different types of structures and is more in line with corresponding distinctions by Descartes, Cudworth or Reid (cf. Mausfeld, 2002, Appendix).

physical energy into neural codes and their subsequent transformations into codes that are 'readable' by and fulfil the needs of the perceptual system¹⁶, the perceptual system contains, as part of our biological endowment, the exceedingly rich perceptual vocabulary in terms of which we perceive the 'external world' (whose relevant aspects do not only pertain to physical and biological aspects but also to mental states of others). Furthermore the perceptual system provides the computational means to make these perceptual concepts accessible to higher-order cognitive systems, where meanings are assigned in terms of 'external world' properties. The sensory system interfaces directly with the motorial system (this interface is evolutionarily an old one) as well as with the perceptual system, whereas the perceptual system interfaces with the motorial system and higher-order cognitive systems. Since the sensory system provides, in terms of its physico-geometrical vocabulary, the cues that the perceptual system exploits in terms of its conceptual structure, issues of cue integration directly refer to the structure of the interface between the sensory system and the perceptual system. In contrast, issues of conjoint representations refer to structural properties of the perceptual system itself.

Although we are still far from having a clear theoretical picture about the kind of primitives that underlie perceptual representations, primitives such as 'surfaces', 'objects'¹⁷ or - as temporal analogues to 'objects' - 'events' (to be understood as *internal*, and not as physical concepts) seem to be among the fundamental pillars of the internal representational structure of perception. These primitives determine the data format, as it were, of internal coding. Each primitive has its own proprietary types of parameters, relations and transformations that govern its relation to other primitives. The data structure for the internal representational primitive 'surface'¹⁸, for instance, can be expected to include a set of free parameters, which refer to attributes such as 'colour', 'depth', 'tex-

¹⁶ Computational approaches of the kind pioneered by Marr almost exclusively deal, with respect to this distinction as I conceive it, with the sensory system; they have revealed that it has a much richer conceptual structure and greater computational power than previously assumed.

¹⁷ Among representational primitives pertaining to 'objects' are, as corresponding evidence suggests, not only those that pertain to 'physical objects' of various types but also a great variety of specific types that pertain to intentional physical objects or to biological objects.

¹⁸ Again, the internal concept 'surface' is assumed to be entirely determined syntactically, i.e. by its data structure and the kind of transformations and relations that operate on it. It is not, in any meaningful sense, a *representation of* physical surfaces. I use the term 'surface' representation only as a convenient abbreviation for a postulated representation (whose nature we presently only poorly understand), whose properties seem to be conveniently describable, at the metatheoretical level of the scientist, in terms of perceptual achievements that are related to actual surfaces.

ture', 'orientation' etc. (again to be understood as *internal*, and not as physical attributes) and may also include specific primitive relations (which may correspond to e.g. 'junctions' and 'edges' of various sorts, 'concavities' and 'convexities', 'gaps', or 'holes').

The values of the free parameters, which lie in a specific region of the corresponding parameter space, have to be determined by the sensory input (and are probably modulated by factors such as 'attentional weight'). The sensory input thus serves a dual function: firstly, it provides triggering cues for which primitives are to be activated, and thus selects among potential data formats in terms of which input properties are to be exploited. Secondly, it triggers processes that result in a specification of the values of the free parameters of the activated representational primitive. Both aspects have to be dynamically interlocked. On the one hand, values can only be assigned to free parameters once the data format has been determined; on the other hand, the activation of a specific data format requires that the values assigned to the free parameters be in a permissible range and lie in a specific region of the corresponding parameter space (if certain types of parameters belong to more than one representational primitives, their values are likely constrained differently). For example, the wavelength information in a sensory input appears to be exploited by (at least) two different types of representational primitives, which we can tentatively refer to as 'surface' and 'ambient illumination' (again, to be understood as *internal*, *not* as physical concepts). The different data formats to which these primitives give rise both include a free parameter for 'colour'.¹⁹ Accordingly, we have to distinguish different types of 'colours', depending on the particular primitive to which they belong. Colours that are attached to the representational primitive 'ambient illumination' subserve a different function and exhibit different coding properties than colours attached to the representational primitive 'surface' (cf. Mausfeld, 2003). The values of the two different kinds of free parameters, which both contribute to the phenomenal attribute of colour, are likely to be subject to different types of constraints.

Although the properties and interdependencies of the free parameters of representational primitives have to mirror, with respect to the perceptual system as an *entirety*, biologically-relevant structural properties of the external world, empirical evidence strongly suggests that they are co-determined by internal aspects, such as internal func-

¹⁹ More precisely, the two different parameters involved can be regarded as pertaining to the same attribute, if they figure as parameters of the same type in some superordinate structures and computations. Again, a label such as 'colour' serves only as a convenient metatheoretical characterisation of a certain type of parameter.

tional constraints or internal architectural constraints, such as legibility requirements at interfaces. The complex and up-to-now poorly understood interdependencies of free parameters contribute to the fact that representational primitives defy definition in terms of a corresponding physical concept (even in the sense of the latter providing necessary and sufficient conditions for the former); rather, they have their own peculiar and yet-to-be identified relation to the sensory input and may also depend intrinsically on other representational primitives, in a way that cannot simply be derived from considerations of external regularities, however appropriately we have chosen our vocabulary for describing the external world.²⁰

In inquiries into the nature of representational primitives, we can, and, taking a specific subsystem of the organism as the unit of analysis, should actually, avoid any notions of the 'proper' object of perception and the 'true' antecedents of the sensory input, among the infinite set of potential causal antecedents (though such notions are, of course, an indispensable part of both ordinary and metatheoretical discourse). The only physics of the external world that figures in a formal theory of visual perception is the physico-geometric properties of the incoming light array. In terms of these properties, we can completely characterize the relation of representational primitives to the sensory input, and thus their 'proximal semantics', as it were, which can extensionally be understood as the equivalence classes of the physical input situations by which they were triggered. The 'proximal semantics' of the perceptual system is, in other words, defined by its relation to the sensory system. The 'proximal semantics' (as a purely syntactically-defined feature) as well as structural relations among representational primitives are given by design and are thus essentially impervious to change by experience. What is modifiable by experience are the values of certain parameters, the latitude of which is determined in a highly specific way that is proprietary to a structure of perceptual representations. Characteristic examples are provided by Wallach and Karsh (1963), who showed that disparity related parameters can be recalibrated by the kinetic depth effect, when disparity and motion provide inconsistent shape information, and by Atkins, Fiser and Jacobs (2001), who showed - with respect to perceiving depth or 3D-shape from 2D-displays in which disks moved horizontally along the surface of a cylinder and exhibited corresponding gradients of texture elements compression - that the differential weighting in the integration of visual cues is recalibrated by the corresponding

²⁰ Because of these interdependencies of free parameters, attempts to identify the representational primitives of the structure of perception and their 'data structure' by investigating attributes like colour or depths in isolation are doomed to fail (apart from lucky coincidences). They are just as futile as it would be to try to determine a n-dimensional manifold from a random sample of one-dimensional projections.

correlations with haptic cues. In animal ethology, illustrative examples are the mechanisms by which birds learn to sing a song appropriate to their species and region (Marler, 1999), or the learning of the solar ephemeris by bees, as part of a sun compass mechanism (cf. Gallistel, 1998). In these cases, a structure of corresponding representational primitives has to allow for parameters whose values are based on a "calibration or checking procedure to insure that the values of those symbols do in fact accurately represent the values of the real world variables to which they refer." (Gallistel, 1998, p. 10) For instance, a sun compass mechanism "has built into it what is universally true about the sun, no matter where one is on earth: it is somewhere in the east in the morning and somewhere in the west in the afternoon. Learning the solar ephemeris is simply a matter of adjusting the parameters of this universal ephemeris function so as to make it fit the locally observed motion of the sun." (Gallistel, 2000, p. 1183) In the case of the human perceptual system, the ontogenetic plasticity, provided by specific structures of representational primitives, cannot, of course, be understood solely on the basis of physical considerations of this kind.

With respect to the structural interdependencies of the free parameters that are potentially involved in a certain input situation, we can, for the purpose of our discussion, distinguish the case dealing with how different parameters of a specific type of representational primitive are interlocked in a certain situation from the more complex case dealing with how parameters of the same type are interlocked in conjoint representations.

When different aspects of the visual input are exploited by the *same* type of representational primitives, for example 'surface' representations, we can encounter situations involving competing interlocked parameters, say for size and distance²¹, orientation and form, or motion direction and form (which can phenomenally be mirrored in multi-stable or vague percepts). A change in the value of one type of parameter, say for coding depth, can, even in cases of otherwise identical stimulus conditions, require strong changes in other types of parameters, say for coding motion direction or 3D-form. The demonstration by Hornbostel (1922) mentioned above is a particularly striking classical

²¹ Even in cases of physically identical input situations, perceptual properties that have usually been regarded as predominantly mirroring properties of input channels, such as discrimination, critically depend upon the settings made for conjoint parameters. A case in point is the observation, known as the Aubert-Förster phenomenon, that discrimination for objects that subtend the same visual angle is better when the object is perceived as a small one at near distance than when it is perceived as a large one at greater distance.

example showing that a change in parameters for motion direction - and a concomitant change in depth parameters - constrains form parameters in a way that is only compatible with non-rigid transformations of form (see also Wallach & Karsh, 1963; Wallach, Weisz & Adams, 1956). Similar observations have been pervasively made with respect to other attributes (e.g. Schwartz & Sperling, 1983; Doshier, Sperling & Wurst, 1986; Kersten, Bülthoff, Schwartz & Kurtz, 1992). For instance, motion can co-determine colour in various ways (Hoffman, 2003; Nijhawan, 1997), and Nakayama, Shimojo and Ramachandran (1990, p.497) observed that "If perceived transparency is triggered, a number of seemingly more elemental perceptual primitives such as colour, contour, and depth can be radically altered." In many of these cases we do not know yet whether we are dealing with the problem of how the different free parameters of a *single* representational primitive are interlocked or with the problem of how representational primitives of the same (or similar) type are interlocked. As a rough experimental diagnostic, one might conjecture that cases in which small changes in a relevant attribute of the input cause radical changes in other attributes indicate situations in which several representational primitives are involved.

Representations that form a conjoint structure are of particular theoretical interest in the present context. In the case of conjoint representations the same aspects of the visual input are simultaneously exploited by two or more representational primitives of the same type or of different types, whose parameter spaces overlap.

In sufficiently complex perceptual systems with a high degree of representational versatility a given (and sufficiently rich) sensory input is not likely to elicit only a single representational primitive but rather triggers conjoint representations. Conjoint representations require special mechanisms and computational means to handle the interlocked way in which they exploit the same input.

Conjoint Representations as a General Structural Property of our Basic Cognitive Architecture

In picture perception, a physicalistically misconstrued framing of the problem of the dual nature of pictures highlighted an interesting class of phenomena while it at the same time obstructing the way to theoretically deal with these phenomena in a fruitful

way. Once general properties of such phenomena are explored on a sufficient level of abstraction, it becomes obvious that cognitive science teems with corresponding phenomena, which witness, it seems to me, our cognitive capability to simultaneously handle conjoint representations over the same input. First, I will illustrate corresponding phenomena by some examples from visual psychophysics, where one can find a plenitude of corresponding examples that indicate that a given sensory input triggers conjoint representations. Afterwards, I will briefly turn to more complex domains, where I must inevitably resign to point out, in a more or less allusive manner, some structural similarities with respect to the issues under scrutiny. For each of these examples, I will first try to identify potential candidates for representational primitives that are integrated in an interlocked way with respect to the same input, and then provide, in an unsystematic manner, some observations that speak in favour of corresponding candidates.

Depth

We presently do not know very much about the representational primitives underlying representations for space and depth. The available evidence suggests that there are, in addition to 'surfaces', probably several quite different representational primitives whose data format is primarily determined by some depth-related parameters.²² Representational primitives for local or distant 'ambient space'²³, or for dealing with 'flat' spatial situations could be regarded as candidates for these representational primitives. Three observations may suffice to illustrate corresponding issues.

i) In a picture, we can see an object as partly occluded but still intact. This can be interpreted as a case of a 2D surface representation competing with a shallow 3D-object rep-

²² Koenderink (1998, p. 1083) argued "that the notion of a depth map as summary representation of pictorial relief is hardly tenable." He concluded that it is likely "that mental structure contains various (perhaps mutually inconsistent) fragments of data structures and that only the execution of particular tasks may perhaps draw on a variety of them and lead to some degree of coordination."

²³ The Ames room demonstration also suggests that an 'ambient space' representation is triggered according to its own rules and has, in cases in which different combinations of values can be assigned to free parameters, its proprietary 'default interpretations', even if these result in ecologically odd parameter settings for 'object' representations pertaining to objects located in this ambient space (in the Ames room, a person who walks along the wall opposite the observer, which physically recedes in depth from the observer, appears to shrink in size).

resentation. Even a simple figure-ground situation, such as Rubin's vase, exhibits a very flat 3D-appearance; the "perceived depth separation between figure and ground has not been well understood", as Weisstein and Wong (1986, p. 33) rightly noted.

ii) Even when cues carrying a strong internal weight, such as occlusion, are violated in a way that is globally incoherent with an interpretation in terms of a physically distant scene, representational primitives with depth-related free parameters can still be activated and provide an impression of space and depth. Magritte's 1965 painting *Le blanc-seing* (depicting a lady on a horse within an assemblage of trees) may serve as an illustration of this. In this painting, occlusion cues that are globally incompatible are provided by locally switching foreground and background, without, however, entirely blocking a depth interpretation (though the resulting impression is a very peculiar one). This applies even more so for cues that do not seem to carry a strong internal weight, such as *linear* perspective or global 3D incoherence of local depth cues (as in so-called impossible figures). Also in this case, a shallow 3D-object representation may compete with a 2D surface representation.

iii) Often representational primitives of the same type seem to be involved in conjoint representations within different frames of reference, as it were. In picture perception, the representational primitive 'surface' is involved both with respect to the perception of the picture surface plane and with respect to the perception of the scene depicted. In this case, different 'surface' representations compete for the depth information that is provided by the incoming sensory input. In the context of picture perception, the visual system seems to have a preference for assigning the differential depth information that is available in the incoming light array to the scene depicted. If the canvas itself exhibits (suitably chosen) differences in physical depth, the physical depth signal of the medium, i.e. the canvas, tends to be internally assigned to the scene depicted. Consequently, the canvas itself looks flat while the picture undergoes corresponding 'distortions'. Striking examples are provided by the paintings of Hughes (cf. Wade & Hughes, 1999).

Colour and Brightness

As depth-related parameters appear, in a highly tangled way, in multitudinous internal representations, it is difficult to identify, in a specific case, the conjoint representations involved. The situation is less complicated when we are dealing with colour- and

brightness-related parameters. Here, phenomenological observations on the interplay of surfaces and (chromatic) illumination as well as corresponding physical considerations provide a rich source for theoretical conjectures about basic conjoint representations. They suggest that the perceptual attributes 'colour' and 'brightness' are part of the data format of two different but interlocked representational primitives.²⁴ They can figure as free parameters with respect to the representational primitive 'surface' as well as the representational primitive 'ambient illumination'. Both representational primitives thus form a conjoint representation with respect to the free parameters 'colour' and 'brightness'. The corresponding regions in the parameters spaces of these two representational primitives overlap. The visual system then has to provide computational means to deal with sensory inputs that are compatible with different parameter combinations in this joint region.²⁵

The interplay of the two representational primitives involved is phenomenally mirrored in many peculiarities that are characteristic for colour appearances under (chromatic) illumination. Of particular interest among these is what Helmholtz (1867) called seeing two colours "*at the same location* of the visual field one behind the other", and what Bühler (1922) referred to as "locating colours in perceptual space one behind the other" (cf. Fuchs, 1923a). For instance, in a room illuminated by a reddish light, we can 'see' both the colour of the object (e.g. 'white' wall) and the colour of the illumination, though there is, as Katz (1911) observed, a "curious lability of colours under chromatic illumination." Similar observations hold, with respect to brightness, for the appearance of surfaces on which a shadow is cast. I will briefly mention a few other observations that seem to be of relevance for attempting to understand the internal structure underlying colour and brightness perception.

i) The dual nature of colour coding that results from the exploitation of the input by two different kinds of representational primitives is perceptually mirrored in what, since Katz's (1911) groundbreaking work, has been called 'modes of appearance', in particular a 'surface colour' mode and an 'aperture or light colour' mode. This descrip-

²⁴ 'Colour' presumably also figures as a free parameter in a variety of superordinate primitives that pertain to more complex biologically-relevant aspects of the external world, such as those pertaining to 'edible things' or to 'emotional states of others'.

²⁵ Mausfeld and Andres (2002) found evidence that second-order statistics of chromatic codes of the incoming light array differentially modulate, by a specific class of parametrised transformations, the relation of the two kinds of representational primitives involved.

tive taxonomy, which itself is in need of explanation by some deeper principles, has unfortunately often been called upon as an explanation itself, thereby confusing the observation with its explanation.

ii) The phenomenal dissociation of brightness and greyness also suggest different representational primitives in which 'brightness' figures as a parameter. Since Hering, it has been well known that even for achromatic colours at least a bi-dimensional account is necessary, as can be witnessed by appearances such as luminous grey. With respect to painting, the difference between a 'brightish white' and a 'whitish bright' is crucial and has been recognized as such since painters became interested in representing the effects of light (Schöne, 1954, p. 203).

iii) The Mach card or Hering's 'stain vs. shadow' demonstration (*Fleckschattenversuch*) are typical classical phenomena that demonstrate how certain attributes can modulate the relation between different representational primitives that exploit a given sensory input. In Hering's demonstrations slight changes in figural characteristics of the Alberti window, namely masking of the penumbra of a shadow by a dark line, are sufficient to induce a switch to a 'surface' representation that completely exhausts the information related to brightness. This is even the case when the physical construction of the situation - that is light source, shadow-casting object and the process of drawing the boundary - is completely transparent to the subject. The available perceptual and cognitive 'interpretations' are completely overruled by a single geometric characteristic.²⁶

iv) The coding properties pertaining to a representational primitive 'ambient illumination' (or transmission medium) resemble, and are probably related to, coding properties of the 'ground' in figure-ground segmentations (cf. Kaila, 1928).

Surfaces and Objects

Representational primitives of 'surfaces' (cf. Nakayama, He & Shimojo, 1995) and 'objects' seem to be among the pillars of the internal conceptual structure of the perceptual system. Hence, it is likely that they themselves are differentiated into families of corre-

²⁶ For recent demonstrations that bear on these issues see, for instance, Adelson (1992), Knill and Kersten (1991), or Buckley, Frisby and Freeman (1994).

sponding primitives that are intertwined in complex ways. From the many types of observations that pertain to this issue I will, in an unsystematic way, mention only three examples.

i) There are situations where we can simultaneously see two surfaces at the same 'location' of the visual field. For instance, looking out of a train window at dusk, we can simultaneously see a red hat on the hat rack and a green tree at the same location in the window. In psychophysics we can find, as Faul (1997) did with respect to chromatic transparency, transitions between transparent and opaque representations of surfaces, or, as Cavanagh (1987) found, conditions under which surfaces are simultaneously opaque and transparent.

ii) With respect to conjoint 'object' representations striking phenomena can be encountered in cases where representational primitives that refer to different levels within some 'hierarchy of organization' are involved, as in the previously mentioned case of so-called impossible figures. An example of other phenomena that can be interpreted along these line are those, referred to as *object superiority effects*, that show that coding properties, such as threshold, detection or identification performance, with respect to elements of the same local stimulus configuration, depend critically upon whether this element is an essential part of an 'object' representation (Gelb, 1921; Lenk, 1926; Weisstein & Harris, 1974; Gorea & Julesz, 1990). On a more abstract level, the process of reading itself seems to be based on an exploitation of a corresponding capacity.

iii) The most general class of phenomena that are likely to be caused by conjoint representations are those traditionally referred to as phenomena of figure-ground ambiguity. The Gestaltist rightly stressed that the figure-ground organization, which refers to internal, mental aspects, not to aspects of the sensory input, belongs to the most fundamental aspects of perception.²⁷ They also observed that different figure-ground organizations for identical sensory inputs would give rise to strong changes in a variety of

²⁷ Problems of figure-ground segmentations are also a "major obstacle in developing computational theories", as Weisstein and Wong (1986, p.61) noted, because basic elements that are used in standard computational theories for the extraction of surface properties are themselves dependent on figure-ground segmentations. Figure-ground segmentation itself is a most fundamental variable that determines and influences perceptual attributes such as colour or depth. Surfaces that are linked up as 'background' can even survive inconsistent disparity information, as Belhumeur (1996, p. 342) showed by a stereogramm in which we perceive a continuation of a background object behind foreground strips, even if this is not consistent with the actual disparity relations for a part of the background section.

other factors such as thresholds or perceptual attributes (e.g. Fuchs, 1923b; Gelb & Granit, 1923). Within the framework put forth here, a change in the figure-ground organization for the same input is regarded as the phenomenal effect of the working of more basic principles that refer to the structure and interplay of competing conjoint representations.

Examples from Other Areas of Cognitive Science

Properties of the architecture of internal representations that have to do with conjoint representations become even more important when the faculties in question have to fulfil, on the basis of a given sensory input, more complex achievements, such as the perception of emotional expressions, or the ability to impute mental states to self and to others. As not much is presently known about the specific data format of representational primitives underlying visual representations, this situation vastly deteriorates when we turn to other areas of cognitive science, beyond vision and language. I will therefore resign myself to listing a few instances that appear to me to share, with respect to the issue of conjoint representations, interesting structural similarities. If these similarities are not merely superficial but rather are grounded in deeper structural properties of our cognitive architecture - and I think there are good reasons to assume so - their careful disclosure would facilitate the identification of theoretical issues that are of great relevance to cognitive science.

Language and Meaning

Our ability to internally handle conjoint representations appears to be mirrored in various ways in language use. Corresponding conjectures appear to gain some plausibility when we deal with how we linguistically handle what is provided by our perceptual system. Many of the examples from visual psychophysics mentioned above implicitly bear on this issue. A more explicit example is provided by the way in which we can simultaneously handle deictic vs. intrinsic, and object vs. viewer centred frames of reference when we are talking about spatial relations (cf. Levelt, 1984; Jackendoff, 1987).

With respect to language and meaning observations of "conflicting perspectives" abound. Although they are undoubtedly of great theoretical importance, it is not easy to assess whether and to what extent an analysis in terms of conjoint representations is

appropriate in these cases (and, more generally, how to conceive of the relation between the 'internal semantics' of the perceptual system and the lexical semantics in the Chomskyan sense). Since the structural similarities appear to me striking, I will, all the same, briefly mention two of them.

In the semantics of natural languages (in the Chomskyan sense of I-languages) the internal conditions governing the meaning of words can encompass simultaneously mutually exclusive aspects of concreteness and abstractness to which we can refer, in using the word, at the same time. "The notion *book* can be used to refer to something that is simultaneously abstract and concrete, as in the expression *The book that I'm writing will weigh 5 pounds.*" Chomsky, 1995, p. 236) It seems to be a pervading characteristic of natural languages that "a lexical item provides us with a certain range of perspectives for viewing what we take to be the things in the world" (Chomsky, 2000, p. 36) and that "quite typically, words offer conflicting perspectives" (ibid., p. 126). For example, "I can paint the door to the kitchen brown, so it is plainly concrete; but I can walk through the door to the kitchen, switching figure and ground. The baby can finish the bottle and break it, switching contents and container with fixed intended reference." (ibid., p. 128)

These and other observations probably point, or so I believe, to interesting similarities between the structure of the lexicon and the structure of the representational primitives of the perceptual system. In language, lexical items provide us with a certain perspective for viewing what we take to be the things in the world; they are "like filters or lenses, providing ways of looking at things." (ibid., p. 36) The same characterization applies, on this level of abstraction, to the perceptual system. Its fixed set of representational primitives provide, as a perceptual ontology, as it were, a set of concepts or a perceptual vocabulary, by which the signs delivered by the sensory system are exploited in terms of notions such as 'surface', 'physical object', 'intentional object', 'potential actors', 'self', 'other person', or 'event' (with respect to a great variety of different categories and time scales), with their appropriate attributes such as 'colour', 'shape', 'depth', or 'emotional state', and their appropriate relations such as 'causation' or 'intention'. In this regard, the structure of the perceptual system seems, in humans, to resemble more the structure of language (more precisely, the structure of the lexicon of I-

language) than the structure of the sensory system.²⁸ The rich conceptual structure of the perceptual system, which extends far beyond physical aspects of the external world, links the signs provided by the sensory system to the conceptual structure of language and of other cognitive systems. The structure of the lexicon, where "notions like actor, recipient of action, instrument, event, intention, causation and others are pervasive elements of lexical structure, with their specific properties and interrelations." (Chomsky, 2000, p. 62), seems to partly mirror (and extend) the conceptual structure of the perceptual system. The representational primitives of the perceptual system and their structure have, in a cognitive system as complex as ours, to ensure an appropriate fit of data formats at the corresponding interface. The property of the lexicon that its items typically offer *conflicting perspectives* on what we take to be the things in the world probably has its counterpart, so I'm inclined to speculate, in the structural organization of the perceptual system in terms of conjoint representations of its representational primitives.

Another, and even more complex, case in point is the use of allegories in oral or written expositions. Their specific properties, as well as the ranges within which they can be employed, probably also reflects the capacity to handle conjoint representations, which bear, respectively, on the relation of medium and message, as it were. Allegories provide a way of expressing something differently from the literal meanings that are used (*aliud dicitur, aliud demonstratur*). For all relevant elements of the exposition they provide two interpretations at the same time, one literal interpretation (*sensus litteralis*) and one actually intended more abstract interpretation (*sensus allegoricus*). One has to understand both at the same time. The literal meaning serves as a kind of semantic medium to trigger the intended superordinate meaning. Similar considerations apply to allegorical meaning in pictorial art, as illustrated in figures 3 and 4.

²⁸ If, to some interesting extent, this should indeed turn out be the case, comparing the functioning of the perceptual system with language, as notably Descartes and Cudworth did (cf. Mausfeld, 2002, Appendix), would not merely be an illustrative or pedagogical metaphor but rather a *theory-constitutive metaphor*, which invites "to explore the similarities and analogies between features of the primary and secondary subject, including features not yet discovered, or not yet fully understood." (Boyd, 1979, p. 363)

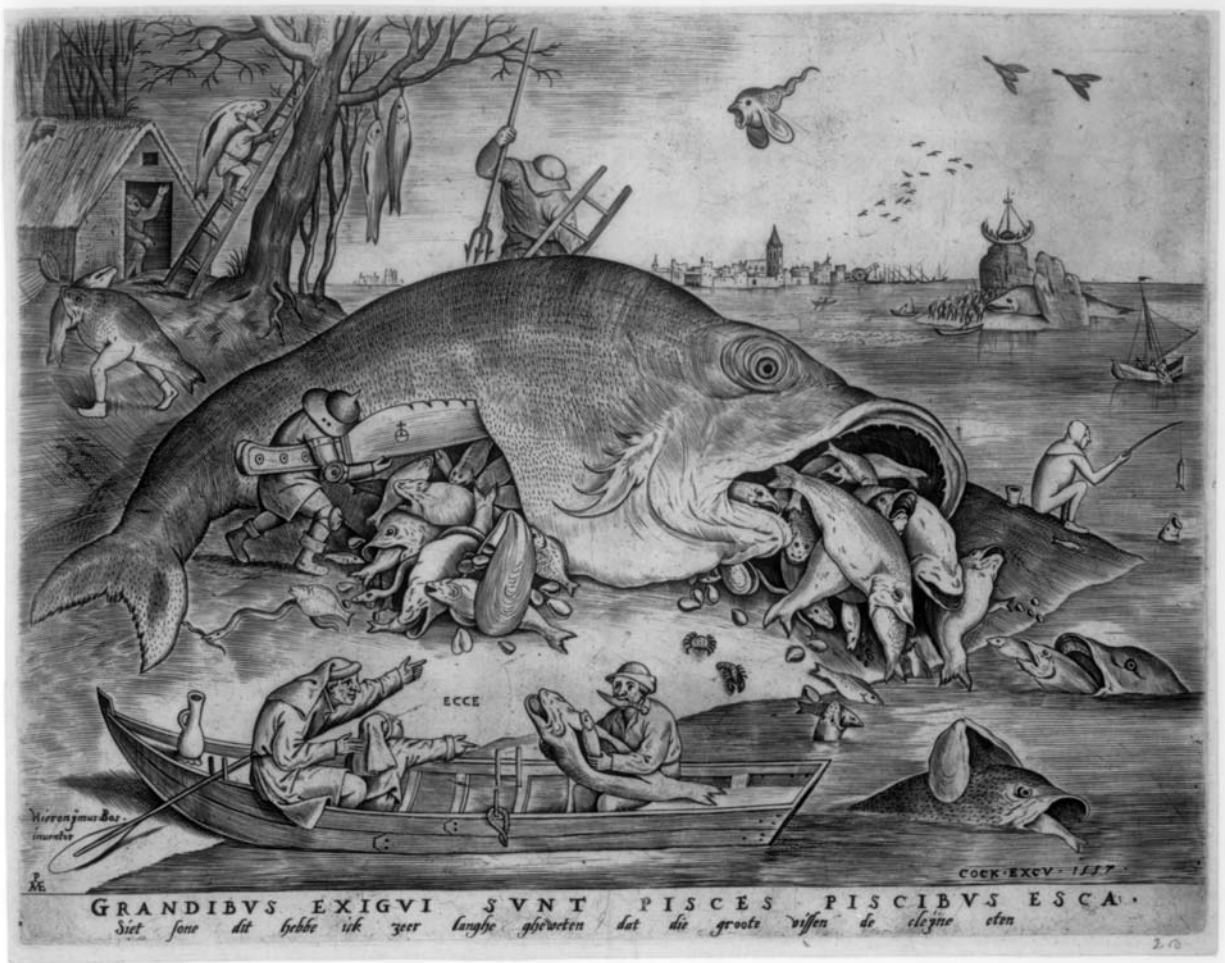


Figure 3: Pieter Brueghel *The Big Fishes Eat the Small Fishes*, 1557, copperplate engraving, FM 1365, Rijksmuseum, Amsterdam

Here, the visual input is exploited by representational primitives that deal with concrete physical objects, as well as by other representational primitives that deal with the perception of social relations and that read through this level of representation and exploit the input in more abstract terms. In the pictures displayed, these abstract terms refer to the 'nature' of social order, as it were.

The allegorical character of figure 4 may appear less obvious, because the objects that figure in the literal interpretation, which refers to people in the Russian concentration camp *Kolyma*, belong to the same category as the ones that figure in the actually in-

tended interpretation, which refers to the threatening of the individual by the terror of the state.



Figure 4: Gerd Arntz *Kolyma*, 1952, linocut, private collection

The pictures displayed demonstrate again that we cannot understand the use of allegories solely in terms of representational primitives. Rather, highly complex interactions of our perceptual faculty and various interpretative faculties are involved about which we presently know, within the framework of cognitive science, next to nothing.

Pretence Play

In pretence play, which is a case of *acting as if*, where the pretender correctly perceives the actual situation, we are also dealing with a case where the same situation is simultaneously exploited by two different representational structures. These structures compete, because, as Leslie (1987, p. 415) put it, "typically the pretense representation con-

tradicts the primary representation." (This contradiction, however, always remains, as Huizinga (1938) noted in another context, constantly fluctuating.) The structural similarity between drama, as a special case of pretence play, and the dual nature of painting was emphasized by Michotte and Polanyi. In the "duplication of space and time that occurs in theatrical representation", Michotte (1960/1991, p. 191f.) noted, "the space of the scene seems to be the space in which the represented events are actually taking, or have taken, place and yet it is also continuous with the space of the theatre itself. Similarly for time also, instants, intervals, and successions for the spectators belong primarily to the events they are watching, but they are left nevertheless in their own present. A further peculiar phenomenon that vividly confirms the unreal character of the representation concerns the way in which an interval, which really lasts usually a matter of minutes or seconds, comes by this process of transportation to have the apparent significance of days, months, or even years." And Polanyi (1970, p. 231) observed that "the painting's self-contradictory flat-depth has its counterpart here in equally paradoxical stage murders and other such stage scenes. .. Art appears to consist, for painting as for drama, in representing a subject within an artificial framework which contradicts its representational aspects."

In infant development spontaneous pretence play emerges at a quite early stage (at about 12 month), and quickly reaches a state, at about the age of 3, where children are able to engage in complex fantasies involving imaginary objects, animals, or people. Furthermore, children are also able to understand the pretence play of other (e.g. Harris and Kavanaugh, 1993).

An explanatory account of pretence play poses, as Leslie (1987) rightly observed, deeper puzzles than reality-oriented play, which responds to an object's actual properties or expresses knowledge about its conventional use. "How is it possible for a child to think about a banana as if it were a telephone, a lump of plastic as if it were alive, etc. If a representational system is developing, how can its semantic relations tolerate distortions in these more or less arbitrary ways. Indeed how is it possible that young children can disregard or distort reality in any way and to any degree at all? Why does pretending not undermine their representational system and bring it crashing down?" (Leslie, 1987, p. 412) Unlike the cases of the dual character of pictures or the dual nature of colour coding, pretence play cannot be understood by referring to an "ability to coordinate two primary representations" (ibid., p. 414) of the same situation. Rather, pretend representations "are in effect not representations of the world but representations of representations" (ibid., p. 417), which makes pretence play a case where a primary represen-

tational structure (which deals with how the situation is actually perceived) competes with a superordinate representation or *metarepresentation* (which deals with what the pretence is).

Imputing Mental States to Others and Perspective Taking

A perceptual system by definition serves to couple the organism to biologically relevant aspects of the external world. For an organism with a mental structure as rich as ours the relevant aspects of the 'external world' do not only pertain to physical and biological aspects but also to the mental states of others. Part of the world as we conceive it are not only objects and surfaces with their perceptual attributes but also emotional states and intentions of others. From an ethological perspective one can reasonably expect that there is, with respect to the architecture and functioning of the perceptual system, no fundamental difference between perceiving aspects of the physical world and aspects of the mental states of others. In either case the sensory input serves as a sign for biologically relevant aspects of the external world that elicits internal representations on the basis of given representational primitives.

Evidently and not unexpectedly, the capability to mentally interact with others is part of the new-born's biological endowment that quickly matures to a state where the child can impute mental states to oneself and to others. The ability to mentally interact with others rests on representational primitives (whose nature is still at the boundary of scientific elucidation) that have their proprietary ways of exploiting the sensory input. It is an essential characteristic of the way these primitives exploit the sensory input that they go 'beyond' those physico-geometrical properties of the sensory input that are exploited by primitives dealing with the physical world. They go beyond what may be called *physical surface characteristics* of the situation encountered and bear on a more abstract construal of this situation. We can see the eyes of a person and the direction of their gaze, and we see a person's face and simultaneously see them being angry or fearful. Figures 5 and 6 may illustrate this, both with respect to a static 2D-representation of a real face and a drawing that depicts a culturally shaped abstraction of a face.



Figure 5: Sergei Eisenstein *Potemkin*, 1926, still from the 'Odessa steps' sequence, New York, The Museum of Modern Art, Film Stills Archive.



Figure 6: Pablo Picasso *Weeping Woman*, 1937, etching, aquatint, and drypoint on paper, Paris, Musée Picasso

From the rich empirical evidence that is available I will only mention one experiment on complex imitation behaviour by Meltzoff (1995). Meltzoff investigated, in a suitably constructed and controlled experimental setting, whether infants ages 18 months interpret "behavior in purely physical terms or whether they too read through the literal body movements to the underlying goal or intention of the act." (ibid., p. 839) In the

critical test situation infants were confronted with an adult who merely demonstrated the 'intention' to act in a certain way, using entirely unfamiliar objects, but never fulfilled this intention. He tried but failed to perform a specific target act, so the end state was never reached and thus remained unobserved by the child. For instance, for the object pair that consisted of a horizontal prong that protruded from a grey plastic screen and a nylon loop, the experimenter "picked up the loop, but as he approached the prong, he released it inappropriately so that it 'accidentally' dropped to the table surface each time. First, the loop was released slightly too far to the left, then too far to the right, and finally too low, where it fell to the table directly below the prong. The goal state of draping the nylon loop over the prong was not demonstrated." (ibid., p. 841) The recorded responses of the infants, namely the number of children who produced the target act, capitalized on toddler's natural tendency to pick up behaviour from adults and to re-enact and imitate what they see. Interestingly, "infants were as likely to perform the target act after seeing the adult 'trying' as they were after seeing the real demonstration of the behavior itself." (ibid., p. 845) They did not re-enact what the adult literally did, but rather what he intended to do (they did not produce the target acts when the physics of the situation, i.e. the movements that are traced in space, were performed by an inanimate device). The type of interaction exemplified in this experiment gives rise to a situation that triggers representational primitives that deal with mental interactions and with the perception of mental states of other. In such situations a kind of *reading-through* with respect to the physical surface characteristics is made possible, which allows the system to organize these characteristics in terms of more abstract mental representations.

A similar reading-through with respect to physical surface characteristics underlies mirror self-recognition. This is an achievement that can only occur when corresponding representational primitives for a self-representation are available, which are interlocked with those that deal with the physical surface situation. Whereas most monkey species under most conditions do not show mirror self-recognition (Tomasello & Call, 1997, p. 336), in humans the relevant representational primitives underlying mirror self-recognition have matured by the age of about 20 months.

There are many other cases of highly complex mental achievements, such as the development of the appearance-reality distinction in young children (e.g. Flavell, Flavell & Green, 1983), whose structural properties probably can abstractly be described in terms of (unknown) conjoint representations, which also may compete for the same input. An important class of representational primitives that have to be assumed as an internal

representational skeleton for perception are also those that deal with dynamic situations or temporally organized events with respect to various types of 'objects' (e.g. Zacks & Tversky, 2001). With respect to picture perception this can be illustrated by the etching displayed in figure 7, where the scene depicted is perceived as a single moment within a sequence of events, and the time slices *not* depicted are as important for what is perceived as the one that is depicted.



Figure 7: Enrico Baj *I funerali dell'anarchico Pinelli*, 1972, etching and aquatint, private collection. It depicts the 'defenestration' of the Italian anarchist Guiseppe Pinelli from the Milan police headquarters on the 15th of December 1969 after the bomb attack by right-wing extremists at the Piazza Fontana.

Structural similarities between these cases appear to suggest that conjoint representations and corresponding transformational structures for properly handling them internally are a fundamental property of our perceptual system. On higher levels of the cognitive system, this property may have its counterpart in our pervading capacity to simultaneously take conflicting perspectives in "looking at things and thinking about the products of our minds." (Chomsky, 2000, p. 36) While an attempt to fit the pieces mentioned above into the theme of conjoint representations is, inevitably, already highly speculative, with respect to the perceptual system, our general capacity to handle simultaneous, conflicting perspectives will almost certainly lie beyond the reach of such attempts. As Chomsky (2000, p. 21) put it: "What we take as objects, how we refer

to them and describe them, and the array of properties with which we invest them, depend on their place on a matrix of human actions, interest, and intent in respects that lie far outside the potential range of naturalistic inquiry." With respect to those aspects that we hope are within the reach of naturalistic inquiry we can, in situations of poor theoretical understanding, only entrust ourselves to Helmholtz's (1867) guiding principle "that order and coherence, even if they ground on untenable principles, are to be preferred to the disorder and incoherence of a mere collection of facts."

Whatever the specific nature of the representational primitives of the perceptual system turns out to be, their categorial character necessitates, from a functionalist point of view, additional general mechanisms for handling continuous transitions in the sensory input, as well as for providing, whenever appropriate, smooth transitions between internal categories. Corresponding questions are of particular relevance with respect to conjoint representations. I will therefore briefly address this issue using examples from visual perception.

Vagueness, Smooth Transitions between Representational Primitives, and the Need of a 'Proximal Mode'

The relation of representational primitives to the sensory input has, in an idealized way, been described above in terms of the equivalence classes of physical situations by which they are triggered, or, equivalently, in terms of equivalence classes of output codes of the sensory system. However, such an idealization is evidently inappropriate, because it would result in a cognitive architecture with functionally undesirable properties. Instead, we have to assume, in line with empirical evidence, that the equivalence classes of physical situations by which representational primitives are triggered have 'fuzzy boundaries', which, in general, yield smooth triggering characteristics both with respect to the relation of a single representational primitive to its triggering class of inputs as well as with respect to transitions between representational primitives that exploit the same input. Since triggering a representational primitive is tantamount to exploiting the sensory input (or the output of the sensory system) in terms of a specific data format with a specific set of free parameters, corresponding 'smoothness' requirements apply, as a rule, to the mappings of physical input features to values of the free parameters.

Usually, in a given input situation (which can also include dynamic sequences of inputs), there is a latitude the extent of which is determined by the structure of the joint parameter spaces involved, as to which representational primitives could be triggered and which values could be assigned to their free parameters; a latitude that corresponds to an ambiguity about which of a set of potential external situations could have given rise to the sensory input. By way of illustration, think, with respect to distance and size, of the Ames room, or, with respect to surface colour and illumination colour, of a white wall under reddish illumination and a reddish wall under white illumination that both give rise to the same sensory input. In such cases the visual system often exhibits a preference for some 'default interpretations'. These preferences can be expected to partly mirror different probabilities by which a certain sensory input can be caused, under 'normal' ecological conditions, by different external scenes. However, such ecological probabilities do not solely or even predominantly determine 'default interpretations', as the cases of the Ames room and the Hornbostel demonstration illustrate. Rather, internal constraints that result from various kinds of stability requirements are, in cases where different combination of values can be assigned to the free parameters, likely to play a crucial role in singling out 'default interpretations'. Global stability of superordinate representations could be maintained by a strategy, with respect to choices between potential values of the free parameters, by which global changes, following small variations in the input (or in the vantage point), in the representational primitives triggered and in the values of their free parameters are, intuitively speaking, kept at a minimum, particularly at the interfaces of the perceptual system with the motorial system and with higher cognitive systems. Such a strategy would protect the system from settling, under 'impoverished' situations, on some definite interpretation that would have to be changed to an entirely different interpretation following a small variation in the input. In input situations whose properties are compatible with various combinations of values of the free parameters (of representational primitives of the same or of different types), transitions between different interpretations often appear to be to some extent receptive to modulations by attentional mechanisms.

Colour perception appears to be a particularly conspicuous case of conjoint representations. Because the same characteristics, with respect to colour or brightness, of a light array reaching the eye can be physically produced in many different ways (e.g. by either a certain interaction of physical surfaces and light sources or, using a slide or a CRT screen, by light sources alone), representational primitives that subservise different distal interpretations, as it were, compete, on the basis of relevant cues, for the same in-

put. These different but interlocked representational primitives, in which 'colour' and 'brightness' figure as free parameters, are referred to above as 'surface' representation and 'ambient illumination' representation. Phenomena related to colour and brightness perception provide rich evidence for the way transitions between representational primitives are handled internally. In the classical literature, corresponding observations were carefully described and their importance was properly acknowledged, despite the fact that a suitable theoretical framework for dealing with them was lacking. Katz, Gelb, Wallach and many others described a plenitude of situations in which "very small changes in external stimulus conditions or in internal modes of perceiving" are accompanied by continuous transitions between conjoint representations, or - as Gelb (1929, p. 600) put it with respect to colour - between internal states that are "of essentially different nature." For example, Turhan (1937, p. 46) observed that, under his experimental conditions, brightness gradients can simultaneously give rise to two incompatible percepts, one of a curved surface (as would result from an 'interpretation' of the sensory input in terms of a specific non-homogeneous illumination) and another one of a slanted flat surface (as would result from an 'interpretation' of the same sensory input in terms of a homogeneous illumination). However, the triggering strength of the sensory input does not suffice to tighten an unambiguous 'interpretation' in terms of either of the representational primitives involved. The internal vagueness with respect to the representational primitives involved is, as Turhan noted, perceptually mirrored in a peculiar impression of perceptual vagueness and indeterminacy.

In colour perception, we can deal with the interplay of the conjoint representations involved - more specifically, with the relation between the corresponding free colour parameters - in terms of the idealized functional goals of illumination invariance and scene invariance of the surface colour at a location of a scene. Because the same sensory input can be compatible with quite different combinations of values of the free parameters (which mirrors the different ways in which the input could have been causally generated) and thus give rise to different functional achievements, the system has to guarantee smooth modulations, under small input variations, of the relations between the representational primitives involved and thus to provide at least a partial compensation between the relevant free parameters. A simple observation that witnesses a corresponding property is provided by the fact that, for instance, a green light (or a greenish ambient illumination) and an olive-green surface, whose colours are yielded by free parameters of different representational primitives, exhibit some phenomenological similarity, although the classes of appearances which these two primitives give rise to could, in principle, have been completely divorced from each other.

An important consequence of the requirement of ensuring smooth transitions between conjoint representations is the existence of what is called a '*proximal mode*' in perception. The existence of a proximal mode is, as Rock (1983, p. 254) noted, "not merely of interest as a phenomenological nicety but rather has important ramifications for a thorough-going theory of perceptual constancy." Evidently, once we have attained the ability to exercise a suitable 'mental attitude', we can perceptually detach certain attributes from their 'frame of reference' as given by a specific representational primitive in which these attributes figure. For instance, a coin lying on the ground at some distance from the observer and being viewed at a slant is perceived as being circular in shape and of its usual size. Still, we can also see it, in the '*proximal mode*', as an elliptical shape of diminutive size. In the same vein, railroad tracks that recede from the observer toward the horizon are perceived as being parallel. Still, we can also see them, in the '*proximal mode*', as converging toward the horizon. Attributes that figure in both types of conjoint representations involved can, apparently, be dissociated from aspects that are proprietary to each of the representations involved. Thus, the existence of a proximal modes helps to protect the system from adopting a behaviour where small continuous changes in the input result in abrupt changes in internal representations.

It is important to note that only those aspects that are necessitated by corresponding continuity considerations are accessible to a '*proximal mode*'. There is no proximal mode in the sense of a measurement-device misconception of perception, or in the sense of the (entirely obscure) notion of a kind of retinal seeing (there is, for instance, no proximal mode for a veridical seeing of isolated elements of so-called geometrical illusions). Rather, what can figure in a '*proximal mode*' is entirely determined by the structure of conjoint representations involved.

In colour perception, for instance, the '*proximal mode*' percept corresponds to those combination of potential values for the free colour parameters of both representations involved that is determined by the internal assumption of a 'canonical' or default situation, which, in this case, would correspond to a spatially homogeneous illumination that does not chromatically deviate from a 'normal' one. The small decontextualized colour patches underlying colorimetry are, with respect to the representational primitives involved, a degenerate situation that is closely related to the '*proximal mode*'. Because such isolated patches proved very useful for investigations into functions of the sensory system that pertain to colour, they often are misleadingly regarded as the building blocks of colour perception. In terms of the representational primitives in-

volved, these isolated patches correspond to in-between stages of *internal* vagueness - which is not to be confused with perceptual vagueness (there is no perceptual vagueness in these cases) -, where the system has not yet been able to settle on a data structure in terms of these primitives.

The percept yielded by the 'proximal mode' is sometimes referred to as the 'local colour quale'. In many situations, one can focus attention on the 'local colour quale' as such, or on colour as a property of surfaces (cf. Arend & Goldstein, 1987); for instance, a spot appearing grey when seen in the first mode of attention may appear as a shadowed part of a white object or a illuminated part of a black one in the second mode. Situations like these, in which it is possible to produce, by slight changes in the mode of attention, transitions where the "surface gains in whiteness to the same extent that the illumination loses brightness" are, as Gelb (1929, p. 600) rightly noted, of "particular theoretical importance." As in picture perception, where we can, with respect to depth, simultaneously have the phenomenal impression of two different types of objects, each of which seems to thrive in its own autonomous spatial framework, we can also, with respect to colour or brightness, encounter situations where we seem to have two mutually incompatible representations at the same time, between which we can, to a certain extent, switch to and fro.

With respect to depth, it is much more difficult than it is with colour to identify representational primitives in which depth figures as a parameter and which are interlocked to form conjoint representations. For biologically crucial internal attributes like 'depth' the corresponding spatial representations are based on a high redundancy from many subsystems in order to guarantee a stable representation even in situations where internal and external conditions deteriorate. In the case of depth, it is particularly difficult to distinguish i) cases of cue integration with respect to the same instance of a representational primitive, say, a specific 'surface' representation, ii) cases in which several representational primitives of the same type are interlocked or compete (e.g. several 'surface' representations in a transparency situation), or iii) cases in which representational primitives of different types are interlocked in conjoint representations. The requirement of smooth corresponding transitions is, however, of importance in each of these cases.

Potential candidates for different and probably conjoint representations in which 'depth' figures as a parameter are, on the one hand, those that deal with the (entirely relative) spatial layout of a scene (e.g. those underlying the kinetic depth effect) and, on

the other hand, those that deal with egocentric distance in a fully-fledged ambient 3D-space. There also seem to be specific mechanisms subserving 'flat' representations by ignoring certain aspects of 3D-structure. For many tasks and operations the availability of a fully-fledged 3D-representation is not necessary or can even be an impediment. Nevertheless the visual system cannot simply discard the corresponding information but has to keep it available internally, because slight changes in the retinal input might require access to 3D-representations. As in the case of colour, this handling of multiple representations can phenomenally be either imperceptible or it can be mirrored in multi-stability or in perceptual vagueness. All of these phenomenal accompaniments can be encountered in picture perception just as much as in other areas of perception. Picture perception is not special, neither with respect to this property nor with respect to other perceptual principles. Like all perceptual tasks that involve human artefacts it rests on and exploits the complex interactions of given perceptual structures and - most notably in the case of non-naturalistic paintings - of various interpretative faculties, whose properties are presently only poorly understood. Among these properties is the ability to phenomenally access the different 'layers' of conjoint representations or to exercise attentional control over them, within the narrow constraints set by the system. Artefacts depend on human intentions and their use is therefore subject to interpretation; this holds for TV screens, microscopes, books or pictures. One has to understand what they were designed for, nevertheless they exploit given capacities. From the perspective of the cognitive sciences, picture perception does not constitute a domain of phenomena that is bound together by some domain-specific explanatory principles. Only when we have become aware that a classification of phenomena in terms of 'picture perception' relies on a pre-theoretical common-sense taxonomy, can phenomena of picture perception prove fruitful for directing our theoretical attention to structural properties of our mental architecture that we otherwise find difficult to notice because they are an all pervading property of the way we are designed.

References

- Adams, P.A. & Haire, M. (1959). The effect of orientation on the reversal of one cube inscribed in another. *American Journal of Psychology*, *72*, 296-299.
- Adelson, E. H. (1993). Perceptual organization and the judgement of brightness. *Science*, *262*, 2042-2044.
- Arend, L., & Goldstein, R. (1987). Simultaneous constancy, lightness and brightness. *Journal of the Optical Society of America A*, *4*, 2281-2285.
- Atkins, J.E., Fiser, J., & Jacobs, R.A. (2001). Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. *Vision Research*, *41*, 449-461.
- Belhumeur, P.N. (1996). A computational theory for binocular stereopsis. In: D.C. Knill & W. Richards (Eds.), *Perception as Bayesian Inference* (pp. 323-364). New York: Cambridge University Press.
- Boyd, R. (1979). Metaphor and theory change: What is "metaphor" a metaphor for? In: A. Ortony (Ed.), *Metaphor and thought* (pp. 356-408). Cambridge: Cambridge University Press.
- Buckley, D., Frisby, J.P., & Freeman, J. (1994). Lightness perception can be affected by surface curvature from stereopsis. *Perception*, *23*, 869-881.
- Bühler, K. (1922). Die Erscheinungsweisen der Farben. In: K. Bühler (Ed.), *Handbuch der Psychologie. I. Teil. Die Struktur der Wahrnehmungen* (pp. 1-201). Jena: Fischer.
- Cavanagh, P. (1987). Reconstructing the third dimension: Interactions between color, texture, motion, binocular disparity, and shape. *Computer Vision, Graphics, and Image Processing*, *37*, 171-195.
- Chomsky, N. (1995). *The Minimalist program*. Cambridge, Mass.: MIT Press.

Chomsky, N. (1996). *Powers and prospects. Reflections on human nature and the social order*. London: Pluto Press

Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.

Deregowski, J.B., Muldrow, E.S., & Muldrow, W.F. (1972). Pictorial recognition in a remote Ethiopian population. *Perception, 1*, 417-425.

Diels, H. (1922). *Die Fragmente der Vorsokratiker*. 4th ed. Vol. I. Berlin: Weidmannsche Buchhandlung.

Dosher, B.A., Sperling, G., & Wurst, S.A. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure, *Vision Research, 26*, 973-990.

Epstein, W. (1968). Modification of the disparity-depth relationship as a result of exposure to conflicting cues. *The American Journal of Psychology, 81*, 189-197.

Farber, J. & Rosinski, R.R. (1978). Geometrical transformations of pictured space. *Perception, 7*, 269-282.

Faul, F. (1997). *Theoretische und experimentelle Untersuchungen chromatischer Determinanten perzeptueller Transparenz*. Dissertation thesis. University of Kiel.

Field, J.V. (1993). Mathematics and the craft of painting: Piero della Francesca and perspective. In: J.V. Field & F.A.J.L. James (Eds.), *Renaissance and revolution* (pp. 73-95). Cambridge: Cambridge University Press.

Flavell, J.H., Flavell, E.R., Green, F.L. (1983). Development of the appearance - reality distinction. *Cognitive Psychology, 15*, 95-120.

Fuchs, W. (1923a). Experimentelle Untersuchungen über das simultane Hintereinandersehen auf derselben Sehrichtung. *Zeitschrift für Psychologie, 91*, 145-235.

Fuchs, W. (1923b). Experimentelle Untersuchungen über die Änderung von Farben unter dem Einfluß von Gestalten ("Angleichungserscheinungen"). *Zeitschrift für Psychologie*, 92, 249-325.

Gallistel, C.R. (1998). Symbolic processes in the brain: the case of insect navigation. In: D. Scarborough & S. Sternberg (Eds.), *Methods, models and conceptual issues. An invitation to cognitive science, Vol. 4.* (pp. 1-51) Cambridge, Mass.: MIT Press.

Gallistel, C.R. (2000). The replacement of general-purpose learning models with adaptively specialized learning modules. In: M.S. Gazzaniga (Ed.), *The cognitive neurosciences, 2nd. ed.* (pp. 1179-1191) Cambridge, Mass.: MIT Press.

Gelb, A. (1921). Grundfragen der Wahrnehmungspsychologie. In: Bühler, K. (Ed.), Bericht über den VII. Kongreß für experimentelle Psychologie (pp. 114-115). Jena: Gustav Fischer.

Gelb, A. & Granit R. (1923). Die Bedeutung von "Figur" und "Grund" für die Farbenschwelle. *Zeitschrift für Psychologie*, 93, 83-118.

Gillam, B.J. (1968). Perception of slant when perspective and stereopsis conflict: Experiments with aniseikonic lenses. *Journal of Experimental Psychology*, 78, 299-305.

Goldsmith, T.H. (1990). Optimization, constraint, and history in the evolution of eyes. *The Quarterly Review of Biology*, 65, 281-322.

Gombrich, E.H. (1982). *The image and the eye.* Oxford: Phaidon Press.

Gorea, A. & Julesz, B. (1990). Context superiority in a detection task with line-element stimuli: a low-level effect. *Perception*, 19, 5-16.

Granit, R. (1924). Die Bedeutung von Figur und Grund für bei unveränderter Schwarz-Induktion bestimmten Helligkeitsschwellen. *Skandinavisches Archiv für Physiologie*, 54, 43-57.

Haber, R.N. (1980). Perceiving space from pictures: A theoretical analysis. In: M.A. Hagen (Ed.), *The perception of pictures, Vol. I* (pp. 3-31). New York: Academic Press.

Hagen, M.A. & Jones, R.K. (1978). Cultural effects on pictorial perception: How many words is one picture really worth? In: R.D. Walk & H.L. Pick, Jr. (Eds.), *Perception and experience* (pp. 171-212). New York: Plenum Press.

Harris, P.L. & Kavanaugh, R.D. (1993). Young children's understanding of pretense. *Monographs of the Society for Research in Child Development*, 58.

Heider, F. (1926). Ding und Medium. *Symposium*, 1, 109-157.

Helmholtz, H.v. (1867). *Handbuch der Physiologischen Optik*. Hamburg: Voss.

Hoffman, D. (2003). Colour and contour from apparent motion. In: R. Mausfeld & D. Heyer (Eds.), *Colour perception: Mind and the physical world*. Oxford: Oxford University Press.

Hornbostel, E.M. von (1922). Über optische Inversion. *Psychologische Forschung*, 1, 130-156.

Huizinga, J. (1938/1986). *Homo Ludens*. Boston: Beacon Press.

Ittelson, W.H. (1952). *The Ames demonstrations in perception*. Princeton: Princeton University Press.

Jackendoff, R. (1987). *Consciousness and the computational mind*. Cambridge, Mass.: MIT Press.

Kaila, E. (1928). Gegenstandsfarbe und Beleuchtung. *Psychologische Forschung*, 3, 18-59.

Katz, D. (1911). Die Erscheinungsweisen der Farben und ihre Beeinflussung durch die Individuelle Erfahrung. *Zeitschrift für Psychologie, Ergbd.* 7.

Kemp, M. (1990). *The science of art. Optical themes in western art from Brunelleschi to Seurat*. New Haven, CT: Yale University Press.

Kersten, D., Bühlhoff, H.H., Schwartz, B.L. & Kurtz, K.J. (1992). Interaction between transparency and structure from motion. *Neural Computation*, 4, 573-589.

- Knill, D.C. & Kersten, D. (1991). Apparent surface curvature affects lightness perception. *Nature*, 351, 228-230.
- Koenderink, J.J. (1998). Pictorial relief. *Philosophical Transactions of the Royal Society London, A 356*, 1071-1086.
- Koenderink, J.J., van Doorn, A.J., Kappers, A.M.L. (1994). On so-called paradoxical monocular stereoscopy. *Perception*, 23, 583-594.
- Koffka, K. (1953). *Principles of Gestalt psychology*. New York: Harcourt, Brace & World.
- Kubovy, M. (1986). *The Psychology of perspective and renaissance art*. Cambridge: University Press.
- Landy, M.S., Maloney, L.T., Johnston, E.B. & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, 35, 389-412.
- Lenk, E. (1926). Über die optische Auffassung geometrisch-regelmässiger Gestalten. In: F. Krueger (Ed.), *Neue Psychologische Studien, Bd. 1*, (pp. 577-612), München: Beck'sche Verlagsbuchhandlung.
- Leslie, A.M. (1987). Pretense and representation: The origins of "theory of mind". *Psychological Review*, 94, 412-426.
- Levelt, W.J.M. (1984). Some perceptual limitation in talking about space. In: A.J. van Doorn, W.A. van de Grind, & J.J. Koenderink JJ (Eds.), *Limits in perception* (pp. 323-358). Utrecht: VNU Science Press.
- Lindberg, D.C. (1976). *Theories of vision form Al-Kindi to Kepler*. Chicago: University of Chicago Press.
- Marler, P. (1999). On innateness: Are sparrow songs 'learned' or 'innate'? In: M.D. Hauser & M. Konishi (Eds.), *The design of animal communication* (pp. 293-318). Cambridge, Mass.: MIT Press.
- Mauries, P. (1997). *Le Trompe L'oeil*. Paris: Gallimard.

Mausfeld, R. (2002). The physicalistic trap in perception. In: D. Heyer & R. Mausfeld (Eds.), *Perception and the physical world*. Chichester: Wiley.

Mausfeld, R. (2003): 'Colour' as part of the format of two different perceptual primitives: The dual coding of colour. In: R. Mausfeld & D. Heyer (Eds.), *Colour perception: Mind and the Physical World*. Oxford: Oxford University Press.

Mausfeld, R. & Andres, J. (2002). Second order statistics of colour codes modulate transformations that effectuate varying degrees of scene invariance and illumination invariance, *Perception*, 31, 209-224.

Meltzoff, A.N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31, 838-850.

Michotte, A. (1948/1991). L'énigme psychologique de la perspective dans le dessin linéaire. *Bulletin de la Classe des Lettres de l'Académie Royale de Belgique*, 34, 268-288. (The psychological enigma of perspective in outline pictures, in: G. Thinès, A. Costall & G. Butterworth (eds.) (1991), *Michotte's experimental phenomenology of perception*, Hillsdale, NJ: Erlbaum.)

Michotte, A. (1954/1991). Autobiographie. *Psychologica Belgica*, 1, 190-217. (Autobiography, in: G. Thinès, A. Costall & G. Butterworth (eds.) (1991), *Michotte's experimental phenomenology of perception*, Hillsdale, NJ: Erlbaum.)

Michotte (1960/1991). Le réel et l'irréel dans l'image. *Bulletin de la Classe des Lettres de l'Académie Royale de Belgique*, 46, 330-344. (The real and the unreal in the image. in: G. Thinès, A. Costall & G. Butterworth (eds.) (1991), *Michotte's experimental phenomenology of perception*, Hillsdale, NJ: Erlbaum.)

Nakayama, K., He, Z.J. & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In: S.M. Kosslyn & D.N. Osherson (Eds.), *Visual cognition. An invitation to cognitive sciences. Vol 2.* (pp. 1-70). Cambridge, Mass.: MIT Press.

Nakayama, K., Shimojo, S. & Ramachandran, V. S. (1990). Transparency: relation to depth, subjective contours, luminance, and neon color spreading. *Perception*, 19, 497-513.

- Nijhawan, R. (1997). Visual decomposition of colour through motion extrapolation. *Nature*, 386, 66-69.
- Pirenne, M.H. (1970). *Optics, painting & photography*. Cambridge: Cambridge University Press.
- Polanyi, M. (1970). What is a painting? *The British Journal of Aesthetics*, 10, 225-236.
- Prazdny, K. (1986). Three-dimensional structure from long-range apparent motion. *Perception*, 15, 619-625.
- Rock, I. (1983). *The Logic of perception*. Cambridge, Mass.: MIT Press.
- Rogers, B.J. & Collett, T.S. (1989). The appearance of surfaces specified by motion parallax and binocular disparity. *The Quarterly Journal of Experimental Psychology*, 41A, 697-717.
- Rogers, S. (1995). Perceiving pictorial space. In: W. Epstein & S. Rogers (Eds.), *Perception of space and motion* (pp. 119-163). San Diego: Academic Press.
- Schöne, W. (1954). *Über das Licht in der Malerei*. Berlin: Gebr. Mann.
- Schriever, W. (1925). Experimentelle Studien über stereoskopisches Sehen. *Zeitschrift für Psychologie*, 96, 113-170.
- Schwartz, B.J. & Sperling, G. (1983). Luminance controls the perceived 3-D structure of dynamic 2-D displays. *Bulletin of the Psychonomic Society*, 21, 456-458.
- Stevens, K. A. & Brookes, A. (1988). Integrating stereopsis with monocular interpretations of planar surfaces. *Vision Research*, 28, 371-386.
- Stevens, K. A., Lees, M., and Brookes, A. (1991). Combining binocular and monocular curvature features. *Perception*, 20, 425-440.
- Tomasello, M. & Call, J. (1997). *Primate cognition*. Oxford: Oxford University Press.

- Trueswell, J.C. & Hayhoe, M.M. (1993). Surface segmentation mechanisms and motion perception. *Vision Research*, 33, 313-328.
- Turhan, M. (1937). Über räumliche Wirkungen von Helligkeitsgefällen. *Psychologische Forschung*, 21, 1-49.
- Wade, N.J. & Hughes, P. (1999). Fooling the eyes: trompe l'oeil and reverse perspective. *Perception*, 28, 1115-1119.
- Wallach, H. & Karsh, E.B. (1963). The modification of stereoscopic depth-perception and the kinetic depth-effect. *American Journal of Psychology*, 76, 429-435.
- Wallach, H., Weisz, A., & Adams, P.A. (1956). Circles and derived figures in rotation. *American Journal of Psychology*, 69, 48-59.
- Wehner, R. (1987). 'Matched filters' – neural models of the external world. *Journal of Comparative Physiology A*, 161, 511-531
- Weisstein, N. & Harris, C.S. (1974). Visual detection of line segments: An object superiority effect. *Science*, 186, 752-755.
- Weisstein, N. & Wong, E. (1987). Figure-ground organization and the spatial and temporal responses of the visual system. In: E.C. Schwab & H.C. Nusbaum (Eds.), *Pattern recognition by humans and machines. Vol. 2. Visual perception* (pp. 31-64) Orlando: Academic Press.
- Willats, J. (1997). *Art and representation. New principles in the analysis of pictures*. Princeton NJ, Princeton University Press.
- Yellott, J.I. (1981). Binocular depth inversion. *Scientific American*, 245, 148-159.
- Youngs, W.M. (1976). The influence of perspective and disparity cues on the perception of slant. *Vision Research*, 16, 79-82.
- Zacks, J.M. & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, 127, 3-21

