What Makes the Eye Intelligent?

A Study in Perceptual Schemata

Alf C. Zimmer

Abstract

A schema is defined as a functional unit of perception or cognition consisting of *primitives* bound together by *rules* (e.g. Gestalt laws of form perception) and a set of *admissible transformations*. When schemata are organized hierarchically, then lower-level schemata form the primitives of higher-order schemata, which in turn impose constraints on the admissible transformations of the lower-level schemata. This definition of schemata in perception implies an interaction between direct and computational modes of perception. The empirical consequences of this notion of schemata have been tested in a series of four experiments about visual orientation in natural and artificial architectural settings. Experiment 1 determined the conditions under which the location of a point-of-view of a picture can be identified. Experiment 2 revealed the consequences of non-admissible transformations on the difficulty of determining the viewing point of a picture in a box world. Experiment 3 was intended to test whether the need to share knowledge about the perceived world influences the process of schematization. Experiment 4 identified how far down in the schema hierarchy top-down constraints on admissible transformations are effective. The general results (processing times and systematic errors) support the notion of highly integrated schema hierarchies in perception. Some consequences for the investigation of cultural influences on perception and on the sharing of world knowledge are outlined.

One of the recurrent experiences with the human senses is that they do not behave like the objective instruments used in physics. Instead of "objective registration" their behavior, or systems characteristics can be best described as providing solutions for perceptual puzzles, which usually make sense for us, or to put it differently, which have a high survival value (KOFFKA 1935). The question why this is the case, and where this apparent intelligence is to be located - in the senses themselves (fn 1) or in higher centers - (fn 2) has puzzled philosophers for more than two thousand years and psychologists for more than one hundred years. The recent controversies about direct vs. computational perception (GIBSON 1979; ULLMAN 1980; FODOR & PYLYSHYN 1981) mirror this old debate and the answers provided bear more than merely superficial similarity to century-old questions. The positions taken concerning the senses' intelligence can be roughly classified into two categories:

(i) The apparent intelligence lies in the direct information pick-up of the affordance provided by the environment (the Empiristic or Realist view).

(ii) The computational processing of simple sensory input provides 'intelligent' solutions for the otherwise chaotic informational influx (the Rationalist view).

For the solution or, at least, clarification of this debate it seems to be worthwhile to return to KANT's (1781) notion of schematism, which was an early attempt to reconciliate empiristic and rationalistic views about perception.

KANT (1799, ed. 5, pp. 176 through 186) characterizes his theory of schemata and its function for the reconciliation of the Empiricist and the Rationalist view of perception as follows. The productive imagination ties sensation and cognition by means of the so-called schema:
"In contrast to the usual case of homology between a concept and its subsumed (real world) instances, for the relation between a category and a sensual object a mediating third is necessary, namely the schema. ... *The schema results from the productive imaginations, ... however, it is not a picture but the imagination of a method how to create a picture for a concept.* ... For instance, the schema of substance is the continuity of the real in time; the schema of causality is the regular sequence of real events."

The 'Oxford English Dictionary' (1971, p. 2663) summarizes this view of schemata as:

"...any one of certain forms or rules of the 'productive imagination' through which the understanding is able to apply its 'categories' to the manifold of sense-perception in the process of realizing or experience".

The concept of perceptual schemata plays an important role in both the direct vs. the computational view of perception, even if it is not mentioned explicitly as, for instance, in GIBSON (1979). NEISSER (1976) points out that many results obtained in the framework of GIBSON's make better sense - even in terms of Realism - if the concept of 'schema' is applied to them. From the Rationalist point of view concepts like 'active mechanisms in perception' (RUMELHART & ORTONY 1976) are brought forward for the description of perceptual information processing. These mechanisms are quite often identified with the concept of 'schema'. What is common for these purported applications of schematism is that they are knowledge-governed processes - much in line with BARTLETT's (1936) notion of schemata in memory and motor behavior but definitely narrower than the one originally proposed by KANT. BREGMAN (1977) discusses the problem from a different angle: he contrasts 'descriptions' with 'ideals' and develops PIAGET's concept of 'scheme' into a model about units in perception, behavior, and cognition. The 'ideals' are assumed to be the elements of psychological processes.

CASSIRER (1944) accommodated KANT's definition of a schema to the results of research on perception as done by HELMHOLTZ, HERING, GELB, KATZ, and the Gestalt psychologists. This is done by interpreting these results in the framework of algebraic group theory and KLEIN's program for geometry (fn 3). He suggests the concept of transformation as an integrative concept for the research on perception; this is very similar to the concept of 'invariants' in direct perception. Starting from CASSIRER's point of view I want to clarify the terminology about higher processes in perception. Furthermore, I suggest that the transformational view of schema theory leads to novel empirical approaches to perceptual problems, such as the problems concerning the perception of space. It is hoped that this approach will make possible the integration of the seemingly contradictory empirical results brought forward in defense of direct or computational perception.

CASSIRER (1944, p. 32-35) identifies the Kantian notion of schema with transformational groups in Kleinian geometry and links it to the crucial question of realism in perception: "Psychology dismisses the dogma of strict one-to-one correspondence between physical stimuli and perceptions. It is, on the contrary, the 'transformed' impression, e.g., the impression as modified with respect to the various phenomena of constancy, which is regarded as 'true' impression, since we can on these grounds construct knowledge of reality" (page 35).

The consequences of such a transformational view of perception are discussed in SHEPARD (1981), in PALMER (1982), and in ZIMMER (1981). In this tradition of thought a schema can be defined as consisting of:

(i) *a set of primitives which are not further analyzable in the given context of perception*, e.g. in the context of line drawings of geometrical solids vertices, angels and lines can be regarded as primitives in spite of the fact that certain lines might be virtual lines, for instance, consisting of a linear change in texture (see GREGORY 1973, pages 89 and 90; GOMBRICH 1973, pages 236 and 237). This sheds light on the fact that on a different level in the hierarchy of schemata the lines themselves can be regarded as schemata consisting in turn of different primitives.

(ii) *a set of organizational rules* which can be paralleled to HELMHOLTZ (1896) logic of 'unconscious inferences' in perception. Examples of these unconscious conclusions are the Gestalt rules of perception: Closure, similarity, symmetry, proportionality, common fate, figure-ground distinction, etc.

(iii) *a set of admissible transformations*, that is, transformations which define the invariance class of objects in question (KATZ 1930), for instance, rotation, translation, projection, etc. How these transformations differentiate between two different schemata which are identical in regard to (i) and (ii) becomes apparent in Figure 1. If the transformation of A into B is admissible, then the schema is a 'line drawing of a cube', if on the other hand only rotational transformations in the plane or in three dimensions are admissible as in C_1 or C_2 then the schema is a 'hexagon with spokes'.

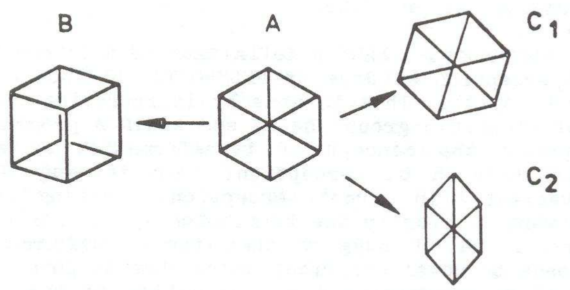

Figure 1 Possible transformations of drawing A: B if A is interpreted as a cube drawing, C_1 or C_2 if A is interpreted as a 'hexagon with spokes'.

From the definition of primitives as determined by the context in question, it becomes apparent that schemata can be assumed to be organized in hierarchies, where lower-level schemata can serve as primitives in higher-order schemata. This kind of organization does not differentiate schema hierarchies from other hierarchical concepts, such as chunking hierarchies. However, the definition of a schema by the appropriate set of admissible transformations can be found in the paintings of ARCIMBOLDO (see e.g. MANDARIARGUES 1977) which are perceived as faces in spite of the fact that they do not consist of of any usual facial parts but of fruit, fish, books, etc. What provides the impression of - strikingly accurate - portraits is the fact that the parts (primitives) obey exactly those transformations which the usual facial parts obey. This second-order isomorphism - that is, of relations between relations (SHEPARD 1981) - is strong enough ot prevail over a total lack of first-order isomorphism.

Figure 2 gives a simplified view of the assumed relations in such a hierarchy of schemata.

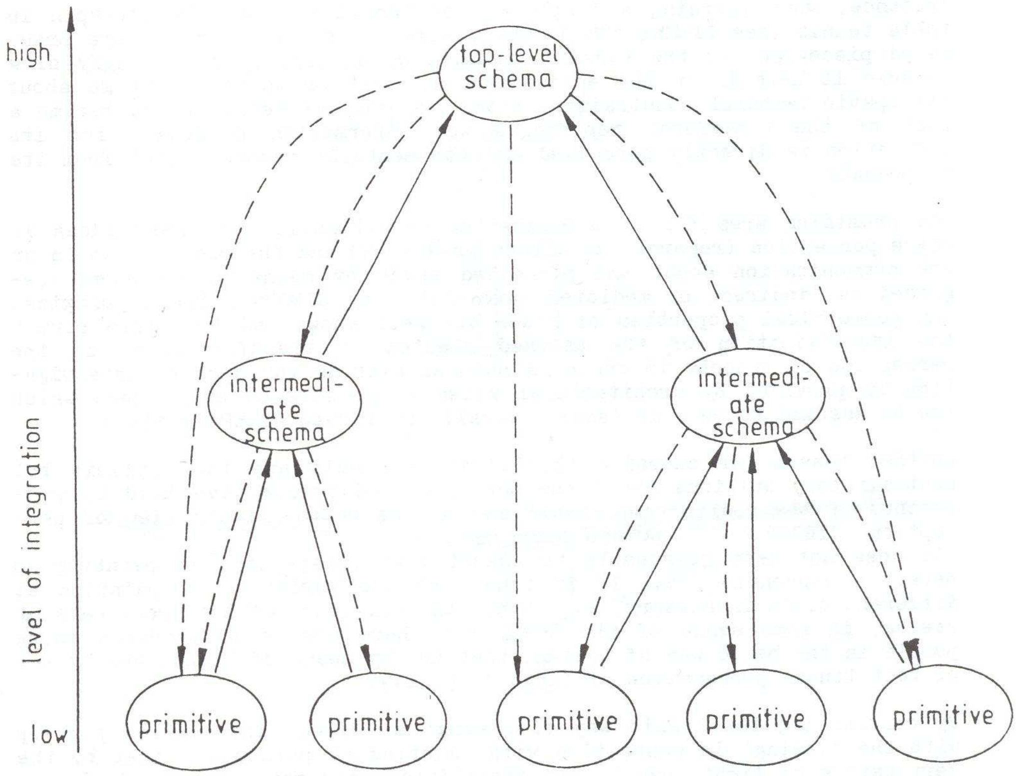

Figure 2 The exertion of constraints in a schema hierarchy. Schemata are indexed according to their level in the hierarchy, the indexing of the primitives is done regarding to the level of the schema into which they are integrated.

The theory of schemata underlying Figure 2 implies that in a way both the proponents of direct and of computational perception are right and wrong at the same time: If one starts from lower-level schemata (e.g. virtual lines consisting of linear changes in textures) given by the affordances of the environment, then on every step upwards it is drawn upon the affordances in apparently the same way as it was done on the lower level. From this point of view, the hierarchical schematic organization is regarded as being an intricate part of the affordance structure (SHAW & TURVEY 1981). On the other hand, starting from very elaborate higher-level schemata and pursuing the hierarchy downwards, conveys the impression that perhaps the amount of implicit knowledge necessary for understanding might decrease but that at the same time the underlying computational processes virtually remain the same. In the approach suggested here it is assumed that constraints due to the evolution of the human species, and constraints due to the necessity for sharing knowledge about the world (see FREYD 1983), lead to a dynamic process of schematic organization. An analogous process can be observed in perceptual development: What requires a cumbersome computation from tacit knowledge at an early learning stage might permit immediate information pick-up later on. For instance, when learning a complex motor behavior (e.g. the top-spin in table tennis (see ZIMMER 1983)) the movements and observations are processed piecewise and the forward planning of actions is accordingly slow because it depends on the activation of overt or tacit knowledge about the spatio-temporal constraints. After the complex behavior has become a part of the behavioral repertoire the information necessary for its regulation is directly perceived and not mentally reconstructed from its components.

One promising area for an experimental investigation of these ideas is space perception (regarded as direct by GIBSON) and the orientation in or the communication about the perceived space by means of pictures (regarded as 'indirect or mediated perception' by GIBSON). The topological and geometrical properties of space are well known and therefore permit the identification of the assumed classes of transformations in the perception of objects in space. A natural setting for such an investigation is provided by architectural sites (e.g. streets or places) which can be assumed to be - at least - locally Euclidean (SHEPARD 1981).

Another reason for choosing this stimulus domain are the explicit but contradictory opinions about the perception of perspective held by proponents of the realist-empiristic and of the rationalistic view of perception. GIBSON (1960) points out:
"It does not seem reasonable to assert that perspective in painting is merely a convention, ... It is true that the varieties of painting at different times in history, ..., prove the existence of different ways of seeing, in some sense of the term. But there are no differences among people in the basic way of seeing, that is, by means of light, and by way of rectilinear propagation of light." (p. 219)

By contrasting the "basic way of seeing" as direct information pick-up with the "seeing" in connection with painting he points out that by the very nature of light, namely its rectilinear propagation the affordances for the perception of perspective are provided. The opposing view, namely that the perception of perspective is an active computational process, it brought forward by GREGORY (1973) in his suggestion that the Inappropriate Scaling Hypothesis explains the MUELLER-LYER illusion: "It (the Inappropriate Scaling Hypothesis) supposes, in the first place, that size and shape constancy are the result of active scaling processes ... In the second place, it supposes that although size scaling can follow apparent distance, even with no change in the retinal image, it can also be set by typical depth cues, even when depth is not seen because it is countered by other information." (p. 75)

Figure 3 gives examples for different techniques for drawing a cube. The drawing techniques differ in the amount how much the higher-level schema 'cube' imposes constraints on transformations on the primitives it consists of. For instance, in the pseudo-perspective technique (Figure 3 b) parallel lines and the length of lines remain unchanged in comparison to the depicted object but on the sides which are not in the fronto-parallel position the right angles are changed consistently into α or $180° - \alpha$.

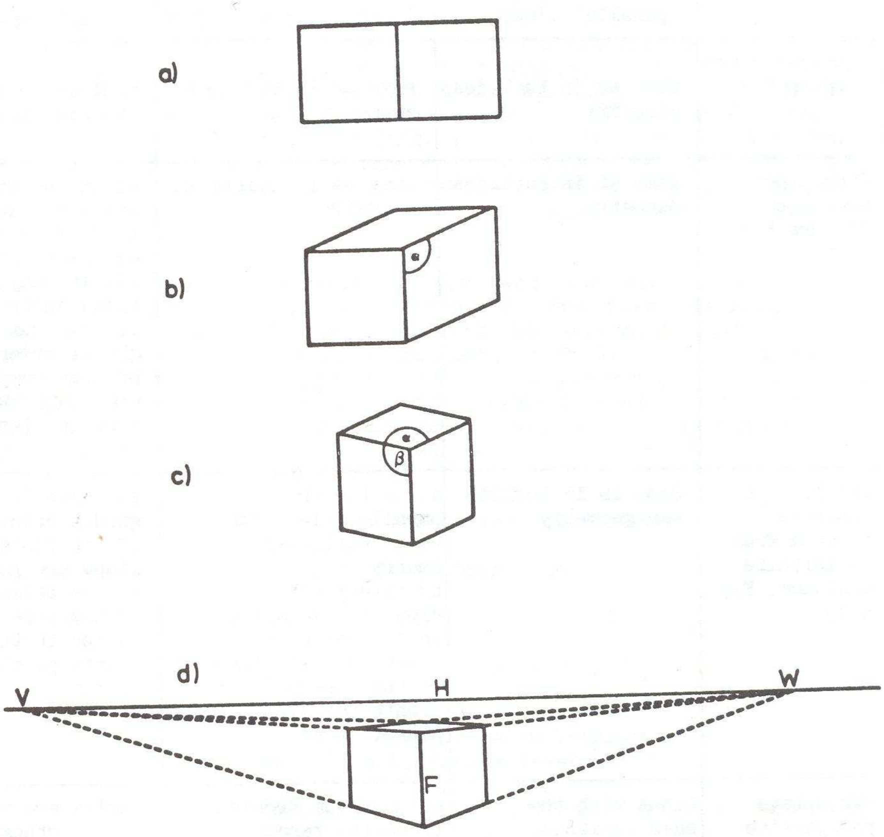

Figure 3 Cube drawings indifferent techniques: a) entirely concept driven, b) pseudo perspective, c) perspective without a finite viewing point, and d) perspective with a finite viewing point.

Table 1 gives the complete list of downward constraints on the admissible transformations of the primitives in the different drawing techniques.

Table 1 Definitions of 'parallel lines', 'lines of equal length', and right angles' in the different drawing techniques of Figure 3.

Primitives

	parallel lines	lines of equal length	right angles
Concept-driven (Figure 3 a)	Same as in Euclidean geometry	same as in Euclidean geometry	same as in Euclidean geometry
Pseudo-perspective (Figure 3 b)	same as in Euclidean Geometry	same as in Euclidean geometry	a) in the fronto-parallel view: same as in Euclidean geometry b) in any other view: Angles equal to the chosen angle of orientation or its complement are regarded as right angles.
Parallel perspective (viewed from an infinite distance, Fig. 3 c)	Same as in Euclidean geometry	a) in the fronto-parallel view: same as in Euclidean geometry b) in any other view: The length of on the angle between the direction of the line and the fronto-parallel plane.	An angle is regarded orthogonal if the connecting lines are parallel to the orientation of the cube in relation to the fronto-parallel plane.
Projective perspective (viewed from a finite distance (Figure 3 d)	Lines with the same vanishing point are regarded as parallel.	a) Only the frontal lines (F) remain main unchanged. b) All other lines are regarded as equally long if their endpoints are marked by the same lines originating from the vanishing points	Angles are regarded as orthogonal a) if they are between lines from a vanishing point and a parallel line, or b) if lines from two different vanishing points meet.

The cube drawings in Figure 3 a) to d) demonstrate in a simplified way the development of drawing techniques in Western art. The most simple one in terms of hierarchical integration is the concept-driven technique as found, for instance, in Carolingian art, when pictures were expected to illustrate what is known and not what is perceived (e.g. the sides of a cube are squares and usually at least two sides of a cube can be seen). Most advanced in terms of hierarchical integration, as seen from the amount of downward constraints, is the perspective technique in the Renaissance period, where the unique point of view of the individual painter and observer was expected to be expressed in a picture. Cube drawings b) and c) lie in between these two extremes (fn 5).

A comparison of the constraints imposed on the primitives by the different techniques reveals that the perspective drawing with a finite point of view (Figure 3 d) exerts the strongest downward influence in the schema hierarchy. At the same time this leads to the most stable depiction of the cube as seen from a definite point of view. On the other hand in the concept-driven drawing (Figure 3 a) no constraints on lower-level schemata exist. The parts known to be constituent for a cube are put together in an additive manner, which does not give rise to any spatial illusion at all.

The set of admissible transformations for each cube drawing which preserves the schema of a cube as constructed from these techniques is identical with the set of transformations in agreement with the restricted definitions of lower level schemata in Table 1. It is possible to regard the cube as the top-most schema in a hierarchy. In this case squares form the next-lower level in the schema hierarchy and so forth. From the constraints on transformations in Table 1 the resulting constraints on squares (Table 2) can be deduced.

Table 2 Admissible transformations for squares in Figure 3.

drawings	admissible transformations
3a	rotation of any multiple of 90°
3b	partial rotation (change of α)
3c	rotation (change of α) view from above or below (change of α) change of orientation (restricted)
3d	rotation (change of position for the vanishing points) change of horizon change of orientation (unrestricted)

What has been demonstrated in this instance can be generalized to the case where cubes and other solids become integrated into schemata like 'buildings' and where buildings in turn are integrated into the schemata of typical architectural settings (e.g. a Roman piazza, a German market place, or a French fortification). The claim is then that the following is true for all construable hierarchies of schemata:

(i) Schemata on the same level are demarcated by a set of integrated lower-level schemata and the admissible transformations (see Figure 1), and

(ii) higher-level schemata impose restrictions on the transformations of lower level-level schemata and thereby on their definition, (See Figure 3 d: in order to be perceived as squares the sides of the cube have be drawn as trapezoids).

The analysis of the relation between the transformational view of schema theory and of spatial perception suggests a test of these assumptions in architectural settings. For instance, a German style market place consists of a well ordered but transformable arrangement of buildings, which in themselves are constructed from geometrical solids, the selection and combination of which underly stylistic and constructional restrictions, and so forth. A further advantage of this stimulus domain is that it allows for strictly controlled laboratory experiments as well as for ecologically richer field experiments. Therefore the relative importance of direct perception in comparison with indirect perception (e.g. of pictures of the setting) can be varied too.

The sequence of experiments starts in a natural setting (a mediaeval place in front of the author's former office in Münster) then proceeds to a manipulable world of card-board box buildings and pictorial representations of it. In all settings there are comparable hierarchies of schemata, which follow the model in Figure 1. Subjects were asked to localize points of view, to communicate spatial knowledge, and to search visually for instances of lower-level schemata embedded in higher-order schemata.

Experiment 1 (fn 5)

The hypothesis underlying this experiment is that each of the higher-order schemata imposes constraints on the lower-level schemata in a predictable way. For instance, the general perspective of the place allows for only a few transformations of the building schemata (there have to be certain kinds of buildings, whereas others are excluded, in other words, they would be "out of place") and so forth.

Method. The experimental site was the place described above, where subjects (10 female and 10 male undergraduates in psychology, attending a lab course in Experimental Psychology) were asked to determine from which position a picture had been taken. The subjects were allowed to take the print and a viewer (having the same angular aperture as the camera) while walking over the place (see Figure 4) searching for the right point of view. When they were convinced that they had found the correct location, the experimenter marked it with chalk. Subjects were paid according to how close their indicated locations were to the target locations. This was done in order to make the motivation for identifying the correct location comparable for all subjects.

The viewing points chosen in taking the pictures can be classified into three categories according to the topological constraints of 'besideness' and 'occlusion' for the buildings in view:

(i) *The stable condition* in which slight (less than 1.0 meter) changes to the right or left of the location of the point of view induces significant changes in the topological constraints. For instance, moving to the right in position (A) leads to an occlusion of buildings 'y' and 'h' and the appearance of building 's'.

(ii) *The semi-stable position* in which slight changes in one direction induce similar changes as in condition (i), whereas even marked changes in the other direction do not induce any change in the topological relations of the buildings. Position (B) is an example for this condition.

(iii) *The instable position* in which marked deviations in both directions do not change the topological relations between the buildings (position C).

For each category 5 points of view were presented.

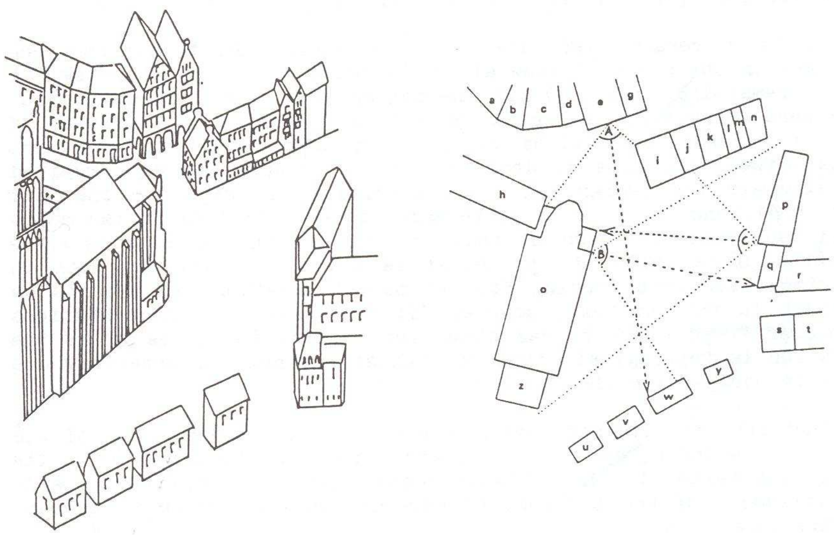

Figure 4 Schematic map of the place in Münster.
A, B, C: points from where pictures have been taken.
----->: indicates the focal direction.
......: indicates the angular aperture.

Results. The distance from each response location to the actual target point was measured. These distances necessarily form a skewed distribution, because there are to be expected many small distances.
Therefore only an ordinal analysis of the data was done. Table 3 the medians and the inter-quartile differences of the distances under the three conditions of topological stability.

Table 3 Distances from the target points (in meters)

Condition	Median	Inter-quartile difference
Stable	2.3	2.1
Semi-stable	4.7	4.8
Instable	5.6	4.3

The data were analyzed by means of the analysis of variances by ranks (KRUSKAL & WALLIS 1952). The overall difference between the conditions is significant ($H = 10.083$; $p < 0.01$). The stable condition differs significantly from both other conditions (MANN-WHITNEY U-test, $z = 2.967$, $p < 0.01$), whereas there is no significant difference between the semi-stable and the unstable condition (U-test, $z = 0.796$, $p >> 0.05$). However, there is a difference between these two conditions if the distances from target in the stable direction of condition (ii) are compared to the distances in condition (iii) ($z = 1.974$, $p < 0.05$). Although, subjects differed markedly in the actual ways they walked while searching for the location the correlations of distances between subjects were all significant (mean correlation: .726, standard deviation of correlations: 0.184, critical value for the Spearman rank correlation, $rho = .52$, $p < 0.05$).

Discussion. These results indicate that the admissible transformations (e.g. changes in the point of view along the horizon with the topological constraints remaining constant) for the schema 'building' are well integrated in perceiving a natural site. Even marked changes in angles or of the size of retinal projections do not induce subjects to recognize deviations between the site as depicted on the picture and as perceived directly (compare the median values of condition (i) with the ones for the other conditions in Table 3). This makes sense from a GIBSONian point of view if one takes into account that "normal" perceiving in such sites usually takes place while the perceiver is moving, as the subjects in this experiment did when looking for the target location. It remains an open question in this context, however, if the internal representations are dynamic as Freyd (1983 b) has shown for various instances or if the representation is topological, that is, robust against perturbations due to changes in perspective when moving horizontally.

The fact that subjects did not take advantage of the substructures of the buildings (e.g. windows, doors, edges, etc.) in order to narrow down the location of the target in the unstable conditions suggests that there is a strong influence of the top-most constraints on the subjects' perception in this task.

The lack of experimental control in this naturalistic setting makes an alternative interpretation of these results possible too, namely that due to the capacity restrictions of short-term memory only the most salient objects and the relations between them were attended to. Under this assumption the only possible strategy for identifying the target position is restricted to the topological information.

If the second interpretation is true, then certain transformations which are non-admissible in the framework of schematic perception of perspective should not influence performance as long as they do not interfere with the critical topological information. The schema-theoretic interpretation, in contrast, would predict any non-admissible transformation to increase the difficulty of the task, since either the schema could not be identified (if there is a violation of admissible transformations which are essential for the schema) *or* the schema could not be integrated into a higher-order schema (if the transformation in question is admissible for the schema in itself but is ruled out due to the constraints exerted by a higher-level schema into which it is usually integrated).

Experiment 2

In order to control for the variables which prohibit an unambiguous interpretation of the results of Experiment 1 the independent variable in this experiment consists of transformations in such a way that the topological constraints between the objects in question were not violated but the perspective constraints were. Any deteriorating effect of these transformations has to be attributed to schematic perception.

Method. The experimental site consisted of 10 car-board boxes (buildings) in different shapes and colors, which formed the standard setting as depicted in Figure 5. The setting was lighted by many diffuse sources in

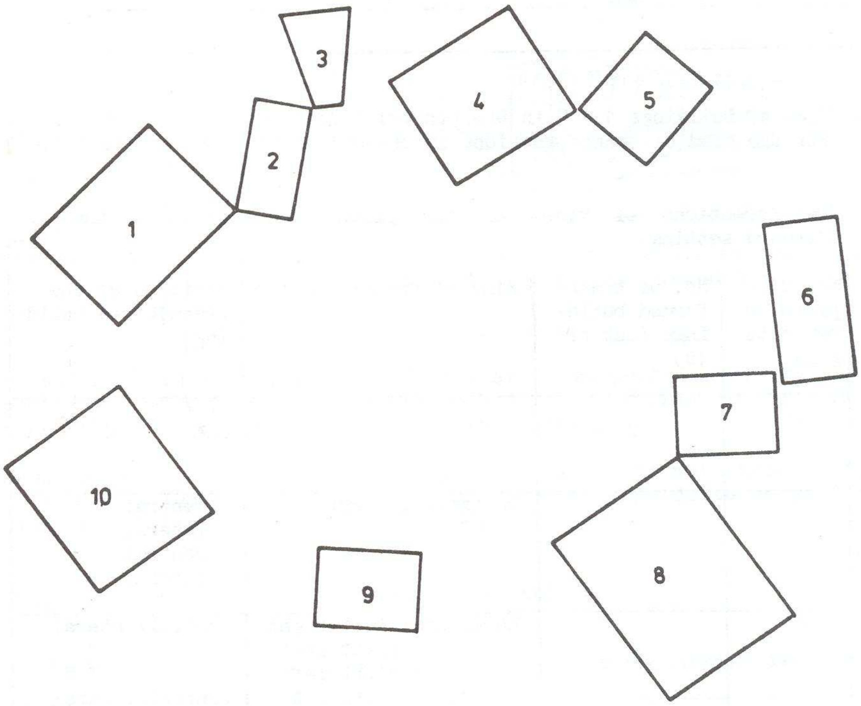

Figure 5 Standard position of the card-box buildings in Experiments 2 and 3.

order to suppress textural and illuminational clues to spatial relations. As in Experiment 1 the subjects' task was to determine the point of view from which the picture had been taken. Each subject (10 male and 10 female undergraduates in psychology) received 28 pictures. In order to vary the independent variable 'number of buildungs which have undergone non-admissible transformations in the pictures', the photographic pictures of one or two of the buildings at different locations in the site were taken from above or below the point of view for the rest of the scene thereby raising or lowering the perspective horizon by ten degrees. The condition 'number of buildings which have undergone admissible transformations in the pictures' was realized by rotating one or two buildings by 10 degrees. In the control condition the photographs depicted the scene veridically (see Figure 6).

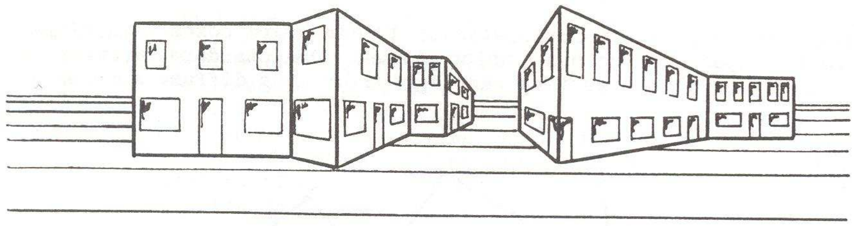

Figure 6 View of buildings 1 - 5 in the control condition.
For the kind of transformations in the different pictures see Table 4.

Table 4 Transformations of views in the pictures compared to the standard setting.

Picture	No. of pictures per category	No. of transformed buildings (out of 10)	Kind of transformation	Position of the transformed buildings
# 1	2	0	CONTROL	
2				
3	4	1	horizontal right	central
4			" "	lateral
5			" "	central
6			" "	lateral
7	9	2	horizontal left/right	central/lateral
8			" right/left	" "
9			" right/left	" "
10			" left/left	central/central
11			" right/right	" "
12			" right/left	" "
13			" left/left	lateral/lateral
14			" right/right	" "
15			" right/left	" "
16	4	1	vertical upwards	central
17			" "	lateral
18			" downwards	central
19			" "	lateral
20	9	2	" up/up	central/lateral
21			" down/down	" "
22			" up/down	" "
23			" up/up	central/central
24			" down/down	" "
25			" up/down	" "
26			" up/up	lateral/lateral
27			" down/down	" "
28			" up/down	" "

Since the horizontal transformations do not violate the topological relations between the card-board boxes (buildings), they form admissible transformations for the single buildings as well as for the general scene. Therefore these transformations should not influence the subjects' performance in identifying the target locations as compared to the control condition. The vertical transformations (raising or lowering the perspective horizon), however, violate the schema-imposed constraints (they are "impossible", see Figure 7) while not changing the topological

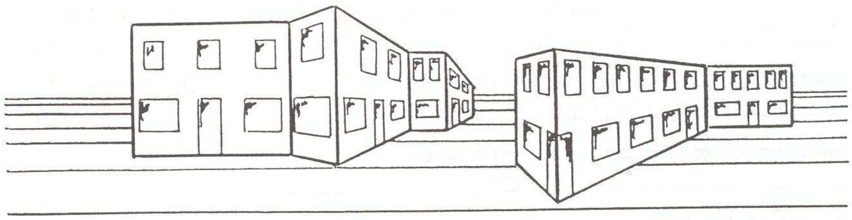

Figure 7 View of buildings 1 - 5 in one of the experimental conditions (condition #18: one central building is viewed against a raised horizon)

relations between the buildings. Therefore subjects' performance is expected to be impaired if their perception is influenced by the schema hierarchy inherent in the architectural setting.

Results. Measuring the precision of localization in meters-off-target led to the puzzling result that there was virtually no differences in precision at all in the three conditions (see Table 5).

Table 5 Precision of localization (in meters)

Condition	Median	Inter-quartile difference
Control	1.7	1.2
Horizontal rotation	1.9	1.5
Shift of horizon	1.9	1.3

The puzzle vanishes if one takes into account the number of subjects who gave up looking for the target after a period of unsuccessful search (Table 6).

Table 6 Number of times when a picture did not lead to the identification of the target point

Condition	Frequency Observed absolute (relative)	Expected absolute (relative)
Control	1 (2.5 %)	2.9 (7.1 %)
Horizontal rotation	5 (1.9 %)	19.0 (46.4 %)
Shift of horizon	35 (13.5 %)	19.0 (46.4 %)

The chi-square value for this distribution is 25.002 (dF = 2; $p < 0.001$). This result is in line with the differences in time necessary to identify the target locations (see Table 7).

Table 7 Identification times (in minutes and seconds)

Conditions	Averages	Standard deviations
Control	2:13	1:17
Horizontal rotation: one building	2:35	0:56
two buildings	2:32	1:22
Shift of horizon: one building	3:15	1:13
two buildings	3:05	1:27

Discussion. From the results in identification times, which can be assumed to measure the difficulty of the task, it becomes apparent that only under the 'shift in horizon' condition is the performance of the subjects impaired in comparison to the control condition (within subject comparison: $t(18) = 2.307$; $p < 0.05$). It has to be kept in mind, however, that these identification times have only been computed for the subjects who succeeded in the task. It can be assumed that, if the subjects were not allowed to quit, the performances under the condition of non-admissible transformations would have shown an even stronger deterioration.

The results from the experiments so far favor the assumption that in the perception of complex spatial scenes two processes occur:

(A) an integration into a general schema (the overall perspective) on the basis of admissible transformations (Experiment 2); and

(B) an extraction of topological informations about the most salient objects in the perceived scene (Experiment 1 and 2).

Both processes are the basis for representations which are used lateron to extract the task-relevant information. The data about the deterioration of performance due to the inconsistencies affecting process (A) allow for three possible models by which these processes are coupled:

(i) *A serial coupling*: (A) precedes (B) and thereby influences the extraction of topological information.

(ii) *A parallel coupling:* the information for both processes is stored redundantly.

(iii) *A parallel and conjunctive coupling:* Only the information which is consistent in (A) and (B) is represented and therefore available for the performance of the task.

Experiment 3

From an evolutionary point of view as taken, for instance, by SHEPARD (1981) who asks for the survival value of perceptual mechanisms, the depicted situation seems to be puzzling at least. In all instances of inconsistency between the processes, the combined representation is more different to the real situation than the sufficient information present in at least one of the processes. One possible solution for this riddle is the following assumption: Not only the veridicality of our perception in relation to the world around us influences the representations (the critical Realist position), but the need to exchange knowledge about the world among members of the same species is decisive too for the representations. If this assumption can be granted the processes (A) and (B) might be related to two different constraints on perception: *veridicality* and *'shareability'* (FREYD 1983 a). Since the experimental conditions in Experiments 1 and 2 do not stress either of these constraints a further experimental task making the exchange of information between subjects topical was conceived in order to clarify the assumed relation between processes (A) and (B).

Method. In this experiment the subjects (the same as in Experiment 1) were randomly assigned to 10 pairs. One number of each pair had to draw a picture of the card-board box scene which should enable his or her partner to identify from which point of view it had been seen. Maps and non-pictorial (verbal or generally symbolic) messages were not permitted. Afterwards the partners had to identify the target location in the same way as in Experiment 2 (control condition). In a second run subjects switched roles.

Results. The overall identification times were about not significantly different from the results for the control condition in Experiment 2 (average value: 2:43 min; $t(38) = 0.674$; $p > 0.25$). This result is surprising, especially from the critical realist point of view, since in the drawings much of the spatial information present in the photographs has been skipped and many details have been distorted (see Figure 8 (a) and (b)). A qualitative description of the drawings illustrates this further:

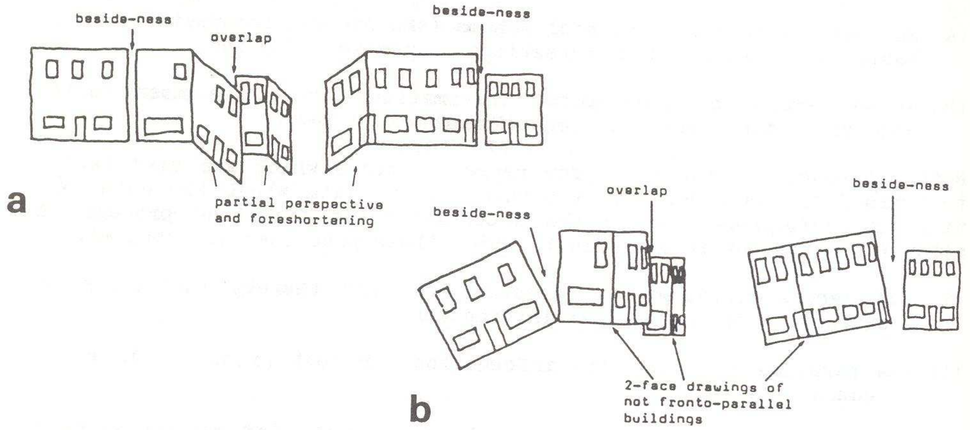

Figure 8 Typical drawings of subjects in Experiment 3:
(a) a partly perspective drawing in a technique similar to the one in Figure 3 b,
(b) an non-perspective (concept-driven) drawing in a technique similar to the one in Figure 3 a.

(i) Only 3 out of 20 subjects attempted to draw the scene with a technique similar to the one in Figure 3 b (see Figure 8 a). This subject and two others used foreshortening, but in all cases the differences in size were smaller than linear perspective would prescribe.

(ii) In the drawings of all subjects the buildings were rotated into a fronto-parallel position in relation to the viewer.

(iii) Fifteen of the drawings contained only topological information about the observed scene; the differences in size were mostly levelled, so that they differed only in shape.

(iv) All drawings were correct in regard to the number, the sequence, and the topological relations (besideness and occlusion) of the buildings.

Discussion. These results indicate that the instruction to draw pictures which can be understood by somebody else leads to sketches mapping nearly entirely the relevant topological information. Only in three cases perspective cues were mapped too: foreshortening in 5 cases and angular transformations as in the pseudo-perspective technique in 3 cases. But the higher amount of information about the spatial setting in these drawings did not induce a better performance in the search task than the entirely topological information in the other 15 drawings.

The fact that it was possible to deduce the relevant information from these impoverished representations of the scene as fast as from the complete information (Experiment 2, control condition) suggests that under this condition subjects processed only the topological structure of the scenes and that this information was sufficient to solve the task.

From a critical realist point of view these results make sense too: As stressed by GIBSON (1979) normal perception does take place embedded in a net of other actions, that is, usually while either the observer or the observed objects are moving. There are two different assumptions underlying approaches to account for this in internal representations:

(i) The internal representation consists of invariants, that is, in the case of spatial scenes the cross-ratios of projective geometry. What is processed in this case are not the objects or the elements they consist of but the cross-ratios between prominent points (CUTTING 1982, Note 1).

(ii) The internal representation is normally dynamic (FREYD 1983 b), that is, the representation captures objects in possible trajectories or transformations they are undergoing while the observer moves.

From an evolutionary point of view the latter assumption is highly plausible, because it explains the factors influencing saliency in perception (KOFFKA 1935) in terms of the information they provide about the dynamics in the perceived world. For instance, the symmetry of a moving object indicates that the observer will be hit by this object. BALL & TRONICK (1971) have shown that even very young children are able to pick up this information. For the situation in the experiments reported here it can be assumed that a double representation exists: one catching the knowledge about the recognized objects and another containing the relational information which makes possible orientation in space.

The result of Experiment 3 exclude the model of purely parallel processes as leading to a higher degree of redundancy, because, if the information were processed in this parallel fashion, then the correct identification of the target could come up in both processes independently. In the long run this would imply both higher accuracy and higher speed in processing. The model of serial processing can be ruled out too, since the perspective integration of the spatial scene is apparently not a precondition for the extraction of topological information. However, the model of *parallel and conjunctive processes* explains, why in perceptual tasks which do not stress the 'shareability' constraints (FREYD 1983 a) the process of the extraction of shareable information and the process spatial schema integration are parallel. It also explains why any inconsistency between these processes influences tasks which depend on these representations.

Experiment 4

In order to test the hypothesis of top-down constraints on the admissible transformations of schemata (see Figure 2), Experiment 4 was designed. It can be assumed that, if an integrative schematizing process of spatial perception like the claimed one exists, then in a purely perceptual task like visual search this process should influence the identification of target objects.

Method. The same subjects as in Experiment 2 were asked to perform the following task: Before every trial one out of four differently decorated window frames was shown. Subjects were asked to memorize its shape and to identify as fast as possible in which building of the subsequent slide this window could be found. Subjects were paid according to their hit rate. Figure 9 shows a typical example of a slide used with possible positions for the target windows on the second floor. The different window frames are depicted below. There was only one target window in

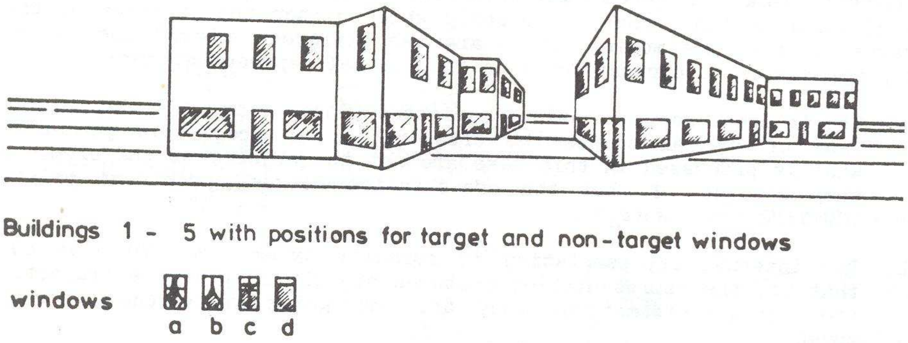

Figure 9 Buildings #1 to #5 with positions in which the target window could appear (2nd floor). Below this example for a slide shown in the experiment there are given the 4 different window types (targets).

each trial. The distractors consisted of the other window frames and non-perspectively drawn specimens of the target window, that is, fronto-parallel drawings at places where the perspective required foreshortening. The latter kind of distractor is especially crucial for the hypothesis in question: If the analysis of the visual scene is done piecewise then this distractor should 'win' over the target because it is physically more similar to the window shown before the trial. If, however, the scene is integrated by the higher-order schema then these distractors should be no more attractive than the other distractors.

The experimental variations of the stimuli were as follows:

(i) *size*, variation between 33 % and 120 % of the size of the window frame shown before each trial,

(ii) *attitude (perspective distortion)*, variations between 0 degrees (fronto-parallel position) and 45 degrees,

(iii) *similarity of the distractors to the target*, this was varied by showing fronto-parallel specimens of the same kind as the target in positions where perspective distortion was required by the spatial scene.

Results. The general hit rate was about 99.7 %. This rate of success was independent of condition (iii), that is, the presence of fronto-parallel distractors had no effect on the success of the identification. Furthermore these distractors did not affect the identification times (3196 ms vs. 3174 ms).

Figure 10 shows the small but reliable effects of size and attitude on the identification times and the interaction between them.

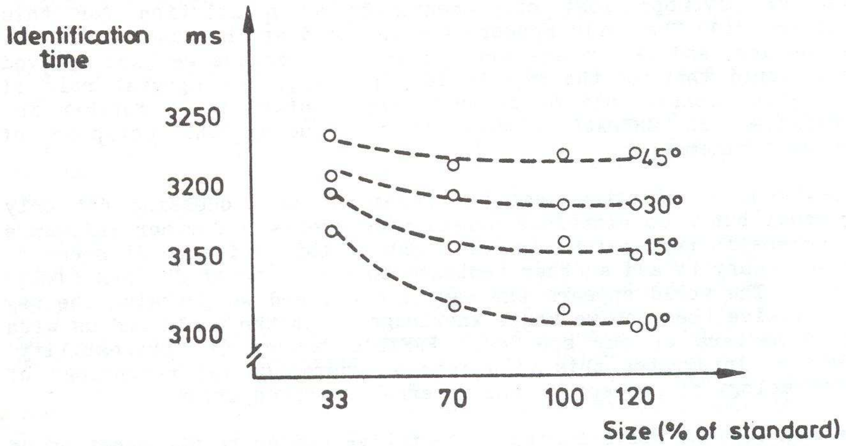

Figure 10 Identification times depending on size and attitude of the target windows.

Discussion. The fact that the fronto-parallel distractors did not impair the subjects performance at all reveals the strong downward constraints of the general spatial schema integrated by perspective on the schemata much lower down in the hierarchy.

The identification times indicate that the subjects need processing time proportional to the amount of admissible transformations necessary in both size and attitude. This result is in line with the results on mental rotation (SHEPARD & METZLER 1971). It can be assumed that during the time necessary for the identification a spatial representation is built up anew from the general perspective clues (that was necessary for every trial since the scenes changed from trial to trial). From this integrated spatial schema the constraints on lower-level schemata (size and attitude of all possible window sites) are derived. Only after this top-down processing is finished the visual search can start. The only way to shortcut this processing would be the choice of the fronto-parallel distractors, but as already mentioned the subjects did not use that strategy.

General Discussion and Conclusion

The experimental results permit some general conclusions and prompt some further hypotheses concerning the role of schemata in cognition. The main conclusions can be summed up as follows:

(i) The model of a schema hierarchy for perception, in which integration and constraints on transformations interact in a predictable way (see Figure 2), is supported by the experimental results on spatial orientation and on visual search.

(ii) Neither ecological perception (GIBSON 1979) nor the computational or tacit-knowledge view of the perceptual process (FODOR & PYLYSHYN 1981; PYLYSHYN 1980) alone seem to be able to account for the experimental data on schematic perception reported here. The mutual influence of environmental constraints and of constraints due to internal processes (Experiment 4) support instead SHEPARD's (1981) view of psychophysical complementarity as a solution for this problem: "(1) The world appears the way it does, because we are the way we are, and (2) we are the way we are, because we have evolved in a world that is the way it is" (p. 332). The special role of ecological constraints on internal representations is further investigated in SHEPARD (1984) by introducing the metaphor of resonant kinematics.

(iii) Experiment 3 indicates that the necessity of processing not only personal but also vicarious experiences exerts a further influence on schematic perception. In the light of this evidence it seems to be necessary to add another sentence to the quote of SHEPARD (1981) above: "The world appears the way it does, and we perceive the way we perceive, because we share knowledge about the world and us with other members of our species." FREYD's theory of 'shareability' (1983 a) integrates this view into a general social psychology of epistemology of perception and internal representation.

(iv) The conclusions above suggest a tentative answer to the question of whether the perspective view of our environment is necessary and general to our species, as GIBSON (1960) seems to imply, or whether it is culturally determined and only found under the environmental and social conditions of a 'carpentered world' (DEREGOWSKI 1973). Whereas immediate pick-up of perspective information is found in Experiment 2 where the situation was environmentally rich, a more computational processing of information seems to underly the behavior of subjects in Experiments 3 and 4 where 2-dimensional representations are predominant for the task. In these cases a strong cultural influence is to be expected. It has to be kept in mind that GIBSON too (1979) claimed that the information pick-up from pictures is indirect.

In the light of conclusion (iii) it becomes possible to interpret the evolution of perspective painting in Western art in terms of 'shareability'. Apparently two polar principles can be found in visual art: Either to depict what one knows, or to depict what one sees. In the art of high-Gothic, influenced by Scholastic philosophy, the maxim of 'transparency' was applied throughout in order to "synthesize all major motifs handed down by different channels." (PANOFSKY 1976, p. 45). This led to highly structured *and* formalized pictures which primarily conveyed the philosophical and theological dogma of that time, whereas the connection to any particular instance of the perceived world was deliberately excluded. The denouncement of the underlying ontological principle "General ideas are real," (fn 6) by the Nominalist philosophers therefore influenced the role of the artist, too. The painter was now looked upon as an 'intuitus', that is, one who paints what he or she sees. This made possible the development of perspective in painting (see the books on perspective by ALBERTI or DÜRER) before the mathematical foundations in projective geometry were laid, because in it the painter's very point of view became crucial for the composition and understanding of the picture.

Intercultural comparisons indicate further societal influences on the importance of perspective in regard to the orientation in space: DAWSON's (1967) results on the perception of perspective cues by people from different African tribes led DEREGOWSKI (1973) to the conjecture that there exists a connection between the social structure, field dependence, and the inability to understand perspective drawings. If this is the case, it becomes quite plausible why the development of perspective painting took place in Italy. There a highly structured, closed mediaeval society broke apart at the same time when DUCCIO and GIOTTO started their experiments with perspective in painting. This example illustrates the mutual influences of society, world knowledge, internal representations, and ways of perceiving as claimed in shareability theory.

The spatial interpretation of two-dimensional perspective drawings is probably only partly due to the Western tradition in art. The fact that we live in a 'carpentered world' is probably more influential and explains why even young children experience various optical illusions which are due to perspective cues: For instance, the Ponzo-, the Mueller-Lyer-, and the horizontal-vertical illusion. Under 'normal', that is, ecologically rich stimulus conditions these illusions play a negligible role (GIBSON 1979), but the related schema-dependent disturbances in veridical perception (Experiment 2) seem to prevail even under such conditions. It can be conjectured that, if the perspective cues are contradictory, this difficulty to integrate information about spatial positions plays a major role in traffic accidents where the correct spatio-temporal orientation is crucial. An example is the preponderance of aircraft accidents on landing strips which are immediately behind a lake or the sea shore. The different temperatures over land and water influence the refraction gradient of the air thereby generating a double-horizon view. This effect is most striking at night, when the visual scene is further impoverished, that is, consisting mostly of dots of light.

Footnotes

The ideas underlying the research reported here have been developed while the author was visiting the Department of Psychology, Stanford University. From the fruitful discussions I have had in Barbara TVERSKY's, Herb CLARK's, and Roger SHEPARD's seminars when reporting these ideas I have very much gained insight in related topics. The theoretical framework owes very much to the long and thorough talks I have had with Jennifer FREYD. I want to thank Gisela REDEKER who has read an earlier version of this paper.

fn 1 Thomas AQUINAS gave this notion the most radical interpretation by counting the senses among the 'virtutes cognoscitivae'.

fn 2 In general, the Rationalist approach to perception can be characrized by the assumption of higher centers governing perception (e.g. PLATO's 'memoria', DESCARTES' 'anima', or PYLYSHYN's (1980) 'computational mind'). These higher centers are regarded as imposing structure upon the physically given thereby molding the perceived world according to the knowledge and rules of the mind.

fn 3 CUTTING (1982) points out that KLEIN's program plays a major role in, what he calls, the reformulation view of direct perception.

fn 4 Even more intermediate states and mixtures of techniques can be found in the history of art. A good example is Bernardo DADDI's Annunciation (about 1340) where he applies techniques b), c), and d) in different parts of the painting, perhaps indicating thereby in which parts he follows iconographic rules and in which parts he follows his own perceiving.

fn 5 Originally, the first experiments had been designed to test subjects' ability to extract geographical information (in this case: the information contained in city maps) from photographic pictures. While there are certain reliable individual differences (e.g. in processing times) it turned out that, more important, there emerged consistent strategies of spatial orientation which ruled the behavior of *all* subjects.

fn 6 The underlying epistemological dogma was therefore called the Realist position - not to be confounded with the Realist point of view referred to earlier in the text -; the counterposition, namely that general ideas are merely 'nomina', is much closer to the modern Realist position.

References

BALL, W.A. & TRONICK, E. (1971) Infant responses to impeding collision: Optical and real. Science, *171*, 818-820.

BARTLETT, F.C. (1932) Remembering. Cambridge: Cambridge University Press.

BREGMAN, A.S. (1977) Perception and behavior as compositions of ideals. Cognitive Psychology, *9*, 250-292.

CASSIRER, E. (1944) The concept of group and the theory of perception. Philosophy and Phenomenological Research, *5*, 1-36.

CUTTING, J.E. (1982) Two ecological perspectives: GIBSON vs. SHAW and TURVEY. American Journal of Psychology, *95*, 199-222.

DAWSON, J.L.M. (1967) Cultural and physiological influences upon spatial-perceptual processes in West Africa. International Journal of Psychology, *2*, 115-128.

DEREGOWSKI, J.B. (1973) Illusion and culture. In: R.L. GREGORY & E.H. GOMBRICH (Eds.) Illusions in Nature and Art. New York: Charles Scribner's Sons.

FODOR, J.A. & PYLYSHYN, Z.W. (1981) How direct is visual perception? Some reflections on GIBSON's "ecological approach". Cognition, *9*, 139-196.

FREYD, J.J. (1983 a) Shareability: The social psychology of epistemology. Cognitive Science, *7*, 191-210.

FREYD, J.J. (1983 b) Dynamic Mental Representations and Apparent Accelerated Motion. Stanford: Doctoral dissertation. GIBSON, J.J. (1960) Pictures, perspective, and perception. Daedalus, *89*, 216-227.

GIBSON, J.J. (1979) The Ecological Approach to Visual Perception. Boston: Houghton & Mifflin.

GOMBRICH, E.H. (1973) Illusion and art. In: GREGORY, R.L. & E.H. GOMBRICH (Eds.) Illusion in Nature and Art. New York: Charles Scribner's Sons.

GREGORY, R.L. (1970) The Intelligent Eye. New York: McGraw-Hill.

GREGORY, R.L. (1973) The confounded eye. In: GREGORY, R.L. & E.H. GOMBRICH (Eds.) Illusion in Nature and Art. New York: Charles Scribner's Sons.

HELMHOLTZ, H. von (1896) Handbuch der Physiologischen Optik, Vol. III, 2nd edition. Leipzig: Voss.

KANT, I. (1799, 5. Aufl.) Kritik der reinen Vernunft. Leipzig: Hartknoch.

KATZ, D. (1930) Der Aufbau der Farbenwelt. Leipzig: Duncker & Humblodt.

KRUSKAL, W.H. & WALLIS, W.A. (1952) Use of ranks in one-criterion variance analysis. Journal of the American Statistical Society, *47*, 614-617.

MANDARIARGUES, A.P. de (1978) Arcimboldo the Marvellous. New York: H.N. Adams.

NEISSER, U. (1976) Cognition and Reality. San Francisco: Freeman.

PALMER, S.E. (1982) Symmetry, transformation, and the structure of perceptual systems. In: J. BECK & T. METELLI (Eds.) Representation and Organization in Perception. Hillsdale: Lawrence Erlbaum Ass.

PANOFSKY, E. (1976) Gothic Architecture and Scholasticism. New York: Meridian.

PYLYSHYN, Z.W. (1980) Computation and Cognition: Issues in the foundation of cognitive science. The Behavioral and Brain Sciences, *3*, 111-169.
PYLYSHYN, Z.W. (1981) The imagery debate: Analogue vs. tacit knowledge. Psychological Review, *88*, 16-45.
RUMELHART, D.E. (1980) Schemata - the building blocks of cognition. In: R.J. SPIRO, B.C. BRUCE & W.F. BREWER (Eds.) Theoretical Issues in Reading Comprehension: Perspectives from Cognitive Psychology, Linguistics, Artificial Intelligence, and Education. Hillsdale: Lawrence Erlbaum Ass.
RUMELHARDT, D.E. & ORTONY, A. (1981) The representation of knowledge in memory. In: R.C. ANDERSON, R.J. SPIRO & W.E. MONTAGUE (Eds.) Schooling and the Acquisition of Knowledge. Hillsdale, N.J.: Lawrence Erlbaum Ass.
SHAW, R.E. & TURVEY, M.T. (1981) Coalitions as models for ecosystems: A realist perspective on perceptual organization. In: M. KUBOVY & J. POMERANTZ (Eds.) Perceptual Organization. Hillsdale: Lawrence Erlbaum Ass.
SHEPARD, R.N. (1981) Psychophysical complementarity. In: M. KUBOVY & J.R. POMERANTZ (Eds.) Perceptual Organization. Hillsdale, N.J.: Lawrence Erlbaum Ass.
SHEPARD, R.N. (1984) Ecological constraints on internal representation: Resonant Kinematics of perceiving, imagining, thinking, and dreaming. Psychological Review, *91*, 417-447.
SHEPARD, R.N. & METZLER, F. (1971) Mental rotation of three-dimensional objects. Science, *171*, 701-703.
ULLMAN, S. (1980) Against direct perception. The Behavioral and Brain Sciences, *3*, 373-416.
ZIMMER, A. (1981) Cultural Constraints on models of cognitive representation. In: R. WILENSKY (Ed.) The Proceedings of the Third Annual Conference of the Cognitive Science Society, Berkeley: Department of Computer Sciences.
ZIMMER, A. (1983) Stadien beim Erwerb komplexer Bewegungsmuster. Sportwissenschaft, *13*, 287-299.

Anschrift des Verfassers:

Prof. Dr. Alf Zimmer
Lehrstuhl für Psychologie II
Universität Regensburg
Universitätsstraße 31
8400 Regensburg