

# SRI International

## AN INDUCTIVE APPROACH TO FIGURAL PERCEPTION

Technical Note No. 325

September 19, 1984

BY: Stephen T. Barnard, Senior Computer Scientist

Artificial Intelligence Center  
Computer Science and Technology Division

SRI Project 5355

The work reported herein was supported by the Defense Advanced  
Research Projects Agency under Contract No. MDA903-83-C-0027.



## **Abstract**

The problem of interpreting single images of abstract figures is addressed. It is argued that neither rule-based deductive inference nor model-based matching are satisfactory computational paradigms for this problem. As an alternative, an inductive approach consisting of two parts is presented. The first part involves a scheme, based on differential geometry, for describing the shapes of curves and surfaces, and for generating these descriptions from images. The second part of the approach relies on a criterion for deciding which description, among the candidates allowed by the constraints in the image, is to be preferred. This criterion — minimum entropy — is related to concepts from Gestalt psychology, thermodynamics, and information theory. Several examples are given to illustrate the inductive approach.

## Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction</b>                                     | <b>1</b>  |
| <b>2</b> | <b>Related Work</b>                                     | <b>3</b>  |
| 2.1      | AI . . . . .  | 3         |
| 2.2      | Perceptual Psychology . . . . .                         | 4         |
| <b>3</b> | <b>A Representational Scheme for Figural Perception</b> | <b>5</b>  |
| 3.1      | Geometrical Descriptions . . . . .                      | 6         |
| 3.2      | Generating Hypothetical Descriptions . . . . .          | 9         |
| 3.3      | Levels of Description . . . . .                         | 10        |
| <b>4</b> | <b>Why are Some Interpretations Preferred?</b>          | <b>11</b> |
| 4.1      | A Model of Structure and Information . . . . .          | 12        |
| 4.2      | Some Examples . . . . .                                 | 13        |
| 4.3      | Discussion . . . . .                                    | 15        |
| <b>5</b> | <b>Conclusions</b>                                      | <b>17</b> |

## List of Figures

|    |  |    |
|----|--|----|
| 1  | Wire Room . . . . .                                      | 2  |
| 2  | Bumpy Torus . . . . .                                    | 2  |
| 3  | The Interposition Illusion . . . . .                     | 5  |
| 4  | Kanizsa's Counter Example . . . . .                      | 5  |
| 5  | The Moving Trihedron . . . . .                           | 6  |
| 6  | Local Surface Types . . . . .                            | 8  |
| 7  | Wire-Bead Backprojection . . . . .                       | 9  |
| 8  | Planar Backprojection . . . . .                          | 10 |
| 9  | Entropy under Change in Amplitude and Symmetry . . . . . | 14 |
| 10 | Entropy under Change in Symmetry . . . . .               | 15 |
| 11 | Entropy under Random Perturbation . . . . .              | 16 |
| 12 | Entropy under Backprojection . . . . .                   | 16 |

## 1. Introduction

Images arise when light that encodes structure in the three-dimensional world is projected onto a photosensitive surface. Some of the information in the light is lost, and the remainder is transformed by perspective into a pattern that has a complex and ambiguous formal relationship to the original structure of the world. The human visual system is capable of inverting this relationship, filling in parts that are missing, arranging parts that are seen into sensible combinations, and, in short, composing integrated, consistent descriptions of the world, which are almost never in serious error. Furthermore, these descriptions specify invariant properties of the scene that are independent of the observer (size, shape, etc.), while the information used to construct the descriptions — the image — is highly dependent on the observer's position, orientation, and imaging system.

How is this possible? What kinds of computational strategies, representations, and modes of reasoning are appropriate to this problem, and how can they be implemented and demonstrated on a large class of examples, including images of natural scenes?

Rule-based deductive reasoning — the conventional AI paradigm — does not appear to be a good approach to perception. Because an image does not logically entail any particular interpretation, one cannot cast the problem of perception in a simple deductive model: interpretations are neither true nor false; they are only likely in varying degrees. But our perception at any moment is unambiguous. Furthermore, our perception sometimes jumps to unwarranted conclusions, as we know from many illusions.<sup>1</sup>

The logical basis of perception is induction. As a mode of reasoning, induction is completely different from deduction. While deduction proceeds from the general (axioms) to the particular (propositions), induction proceeds from the particular to the general. Deduction is primarily a matter of proving theorems, while induction is one of recognizing patterns. Deduction is well-understood and more easily automated with computers, which probably explains its popularity in AI research. The mathematical foundations of induction, by contrast, are much less clear. Nevertheless, general principles of inductive reasoning do exist.

It has been postulated that the uniformity and regularity of the world are necessary presuppositions of induction. This is precisely the state of affairs in perception. The underlying reality (the scene) is not logically deducible from the image, but, in most cases, a very good guess can be made by finding the simplest possible interpretation.

Specifically, The problem of **figural perception** is defined as deciding how to assign three-dimensional properties — size, shape, position, orientation, etc. — to initially two-dimensional patterns of data. The patterns of interest vary in their degree of complexity. For example, they might be simply binary contours, such as Figure 1. The sense of realism in even these simple figures compels one to believe that very general perceptual processes apply. A somewhat more complex class of patterns is synthetic intensity images, such as Figure 2, in which a combination of surface, lighting, and projection models produces images that evoke an even more vivid impression of three-dimensional shape.

Figures 1 and 2 are synthetic: they were generated with the techniques of computer graphics [1]. The Bumpy Torus, for example, was created by constructing a smooth, randomized toroidal surface, defining a reflectance function with lambertian and specular com-

---

<sup>1</sup>In a strictly logical sense, perception *always* jumps to unwarranted conclusions.

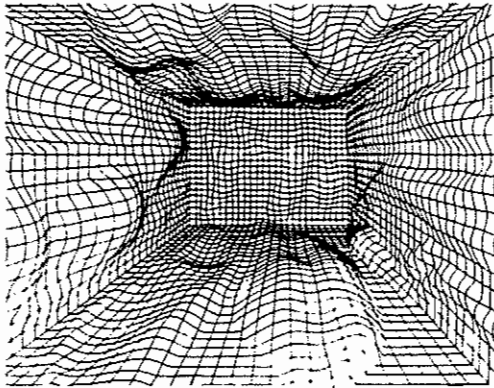


Figure 1: Wire Room

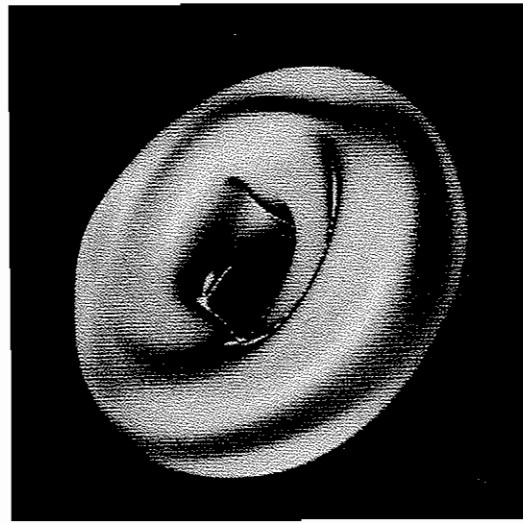


Figure 2: Bumpy Torus

ponents, defining a lighting model and a viewing position, and, finally, centrally projecting the intensities of a very fine mesh of surface points onto a synthetic digital image. A depth buffer was used to handle hidden surface areas. Using synthetic data has two important methodological advantages: (1) the underlying reality is known to arbitrary precision and can easily be used to evaluate interpretations, and (2) variables that are difficult to control in physical imaging, such as lighting and film response, are easily controlled in a synthetic regime. Of course, if a theory of figural interpretation is to have practical importance, it must be applicable to real images. If a computational vision technique works well on very realistic synthetic images, without relying on special conditions that are known *a priori* (such as a specific lighting model), then it will probably work well on comparable real images. If the technique shows improved performance on images that are subjectively more realistic, we can be even more confident that it will be valid for real images.

The physical constraints in the problem of figural perception, while obviously important, are insufficient: infinitely many possible surfaces could have caused these figures, but our perception chooses only one. The thesis behind this paper is that a formal geometrical language, together with general principles of inductive reasoning, can account for at least a large part of the solution to this underdetermined problem. A geometrical language, combined with physical constraints, provides a space of possible three-dimensional descriptions or “explanations” of patterns, and inductive reasoning provides a basis for choosing among them.

The inductive approach to figural perception has two critical elements:

- First, there is a representational scheme, based on vector algebra and differential geometry, that can model the image and all of its possible interpretations. Implicit in this scheme is a process for *generating* interpretations (Section 3).
- Secondly, there is an inductive criterion for preferring certain interpretations over others. This criterion — minimum entropy — is based on a formalism originally developed for statistical mechanics. In the context of figural perception, entropy is

used as a measure of disorder (Section 4).<sup>2</sup>

The approach treats perception as a search for the simplest explanation of a body of data (an image). An interpretation is therefore a re-encoding of an image. Properties and relations that are explicit or easily computed from the image (pixel values, edges, textural properties, etc.) become implicit in the re-encoding and may be at least partially recovered by reprojection. On the other hand, properties and relations that are merely implicit in the image (scene invariants, such as shape, size, orientation, relative position, reflectivity, transparency, etc.) are explicit in the re-encoding. The image is unstructured and lengthy: it contains redundant information. The re-encoding is structured and terse: it contains at least as much information as the image, and usually more, but in a compressed form, with the redundant part removed. Some process not yet fully understood *discovers* redundancy in the image and *exploits* this redundancy to build more concise and well-formed encodings. In practice, it may not be necessary to actually construct a concise encoding, but merely to recognize that one is possible.

It is useful to think of an agent that “decodes” the final interpretation and that has the knowledge and ability of a computer graphics system. The 3D encoding describes the scene in terms of physically meaningful, invariant properties. The agent can decode it, in principle, into a “visualization” of the scene by using an abstract model of projection, a choice of viewpoint and lighting, and specific knowledge of physical principles, such as that an opaque object occludes what is behind it or that a transparent object transmits light. Therefore, while the interpretation contains no less information than the image, it is in a form that makes the important invariant properties explicit, and relegates the ones that depend on viewpoint and lighting to an implicit status.

## 2. Related Work

Two distinctly different schools of research have addressed the problem of figural perception. The artificial intelligence (AI) approach has focused on computer implementations, while the perceptual psychology approach has developed primarily theoretical models. The scientific methods used in the two disciplines are quite different. Vision research in the AI style generally requires precise computational models of perception: if a theory cannot be implemented, it is too vague to be of value. Ultimately, the model should be evaluated on images of real scenes. Vision research in perceptual psychology, by contrast, has sought to explain *human* perception as revealed by illusions, psychophysical experiments, and introspection. Perceptual psychology is by far the older school, and AI has borrowed from it liberally. At the same time, the development of computers has influenced psychologists to pursue information-processing approaches and to embrace concepts originally developed in AI [5].

### 2.1. AI

The deductive approach to figural perception has been explored in the so-called “blocks world” work (see Mackworth [2] for a summary of this research), culminating in Waltz’s fil-

---

<sup>2</sup>While the representational scheme is based on the geometry of curves and surfaces, the reasoning scheme has far broader generality.

tering technique for constraint satisfaction [3], and Kanade's generalization to the Origami world [4]. The results are not encouraging. In addition to the problem of needing a perfect line drawing to begin with, these systems produced only weak interpretations, not including, for example, quantitative estimates of length and orientation. When generalized only slightly, Waltz's filtering scheme led to many more ambiguous interpretations.

Another line of AI research, which is more relevant to the approach described here, has sought *metric* interpretations of images, as opposed to the weaker, merely descriptive interpretations characteristic of the blocks world. The first instance of such an approach was due to Huffman [6], who suggested the concept of **dual space**, later generalized by Mackworth [7] to **gradient space**. Gradient space simply provides a way of representing with two parameters the orientations of planes. Mackworth connected observed features in image space (vertices) with constraints in gradient space (i.e., constraints on the orientations of planes) to disambiguate blocks-world interpretations. Kanade used gradient space to estimate orientations on the basis of symmetry [8]. That is, image figures that exhibit skewed symmetry (because of the distortion introduced by projection) are interpreted as being oriented in a way that is consistent with their true symmetry.

This general approach — identifying an important invariant property in the plane, back-projecting image features to planes of different orientations, and selecting the orientation leading to the most well-formed configuration — has been followed by several researchers. Kender [9] used textural properties, such as the lengths and orientations of line segments; Ikeuchi [10] and Barnard [11] used angles; Witkin [12] sought the planar orientation that had the most uniform distribution of directions of contour tangents; Brady and Yuille [13] maximized the compactness of the backprojected closed contour; and Barnard [11] maximized the uniformity of backprojected curvature. The inductive approach can possibly unify these various criteria into a single principle.

Another area of AI vision research that is relevant to figural perception is the optimal interpolation of surfaces [14], [15], [16], [17]. The mathematical representation of surfaces and the optimization methods used in this work have similarities to the approach described here. The underlying problems are quite different, however. The problem of optimal interpolation is to begin with sparse three-dimensional data (distances and orientations), presumably derived from stereo, shape-from-shading analysis, etc., and to find a continuous surface that best fits the data, while optimizing physical properties of the surface (specifically, potential energy). The problem of figural perception initially provides no three-dimensional information at all, and is not even well-posed in the sense that the interpolation problem is. Furthermore, we choose interpretations according to their simplicity of description, and not according to a physical property.

## 2.2. Perceptual Psychology

A popular approach in perceptual psychology has sought to exploit the efficacy of information theory [18], [19], [20], [21], [22], [23], [24]. Rock calls this the modern version of Gestalt theory ([5], p. 133), because its aim, just like Gestalt, is to explain perception in terms of simplicity. While there is not space here to cover all this work, it will be useful to discuss in some detail a recent approach that has some similarities to the approach presented here.

Buffart, et. al. presented a "coding theory" of perception that was meant to explain



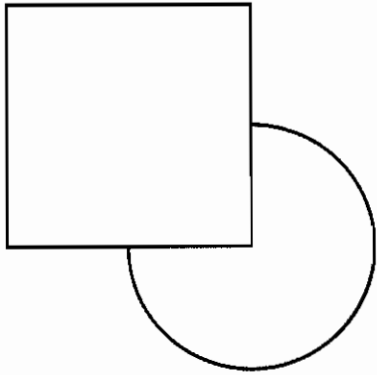


Figure 3: The Interposition Illusion

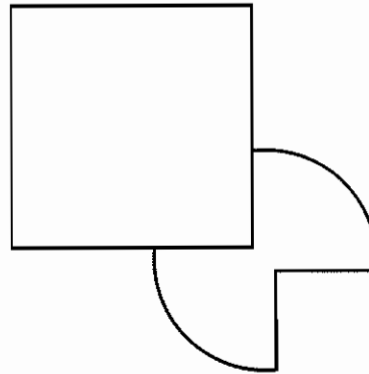


Figure 4: Kanizsa's Counter Example

the interposition illusion [25]. Most observers see the pattern in Figure 3 as a square on top of a circle. Coding theory attempts to explain this by asserting that a description in terms of a square on top of a circle is simpler than any other description that accounts for the figure. The authors proceed to develop a coding scheme for these figures that takes advantage of symmetries and that leads to very concise encodings. The encodings are sentences in a formal language, with the primitives representing sides, angles, circular arcs, and combinational operators. Some context-sensitive elements are included; for example, a side can be extended indefinitely until it encounters another contour. The goodness of an encoding is determined by simply counting the number of symbols it uses.

There are several objections to this theory. First, Kanizsa [26] argues that a pattern such as Figure 4 is a counter example, because the interpretation without interposition is simpler than the one with interposition: the circle with two "bites" taken from it has two axes of symmetry, and should, therefore, be more symmetric, and hence simpler, than the one with only one bite. As will be shown in Section 4.2, this objection is not valid. That a figure has more axes of symmetry than another does not imply it is simpler.

A second, more serious objection to the coding theory is that it depends on an *ad hoc* language, and there is no compelling reason to adopt this language in preference to any other. A third objection is that, even given this particular language, mere symbol counting is not a good way to measure the complexity of an encoding. A fourth objection is that no procedure for actually constructing a minimal encoding is presented. The approach presented below, when considered as an alternative to the coding theory, meets these objections.

### 3. A Representational Scheme for Figural Perception

The view of perception as a computational process of building, testing, and selecting descriptions is arguably the most important contribution of artificial intelligence to perceptual psychology. When faced with the task of actually implementing a computational model of perception, one must deal with representational problems that are otherwise too easily ignored. If perception is description building, what must these descriptions be like? In

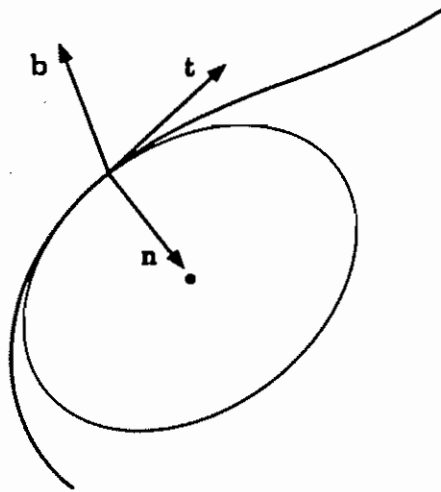


Figure 5: The Moving Trihedron

what kind of language should they be expressed?

### 3.1. Geometrical Descriptions

The problem of figural perception, is to a large extent, a problem of geometrical description. We seek interpretations in terms of geometrical objects: points, curves, and surfaces. The description of the special cases of points, straight lines, and planes is relatively straightforward: these objects can be represented with vector algebra [27]. Much more difficult is the representation of general curves and surfaces.

Differential geometry is the study of geometric figures using the methods of calculus [28]. Three requirements compel us to use the language of differential geometry in our representational scheme:

- If we are to compare descriptions on the basis of simplicity, we must have canonical descriptions. The descriptions must be *unique*.
- The language must be expressive enough to describe the entire range of figural phenomena. It must be *complete*.
- The descriptions should express *intuitive* and *invariant* figural properties.

The form of the invariant properties of curves and surfaces embedded in three-dimensional Euclidean space is completely known for our purposes.

Any curve  $\mathbf{x}(s)$  in  $C^2$  (i.e., any twice-differentiable curve) can be represented with two invariant local properties, curvature  $\kappa$  and torsion  $\tau$ , that are scalar functions of arc length,  $s$ , and that constitute a complete, unique, and invariant representation of the curve. The relationships are described by the Serret-Frenet equations:

$$\begin{aligned} \dot{\mathbf{t}} &= \kappa \mathbf{n} \\ \dot{\mathbf{n}} &= -\kappa \mathbf{t} + \tau \mathbf{b} \\ \dot{\mathbf{b}} &= -\tau \mathbf{n} \end{aligned} \tag{1}$$

where  $t$ ,  $n$ , and  $b$  are, respectively, the tangent, normal, and bi-normal vectors (Figure 5). The dot operator indicates differentiation with respect to arc length. The important point is that a description of a curve in terms of curvature and torsion is independent of the choice of a coordinate system. Barnard and Pentland [29] have studied the interpretation of images of 3D curves with torsion by using local assumptions of maximally uniform curvature and constant torsion.

Using the concepts of differential geometry, a surface  $\mathbf{x}(u, v)$  in  $C^2$  can also be represented with invariant local properties. The relationships analogous to the Serret-Frenet equations are the Gauss-Weingarten equations:

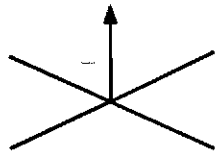
$$\begin{aligned}
 \mathbf{x}_{uu} &= \Gamma_{11}^1 \mathbf{x}_u + \Gamma_{11}^2 \mathbf{x}_v + LN \\
 \mathbf{x}_{uv} &= \Gamma_{12}^1 \mathbf{x}_u + \Gamma_{12}^2 \mathbf{x}_v + MN \\
 \mathbf{x}_{vv} &= \Gamma_{22}^1 \mathbf{x}_u + \Gamma_{22}^2 \mathbf{x}_v + NN \\
 \mathbf{N}_u &= \beta_1^1 \mathbf{x}_u + \beta_1^2 \mathbf{x}_v \\
 \mathbf{N}_v &= \beta_2^1 \mathbf{x}_u + \beta_2^2 \mathbf{x}_v
 \end{aligned} \tag{2}$$

where  $\mathbf{N}$  is the unit normal to the surface, and the subscripts  $u$  and  $v$  indicate partial differentiation. The coefficients  $\Gamma_{ij}^k$ ,  $\beta_i^j$ ,  $L$ ,  $M$ , and  $N$  are determined by the local shape of the surface. The theory of surfaces is much more elaborate than the theory of curves, as a comparison of Equations (1) and (2) suggests.

To develop an intuitive understanding of the power of the theory, consider the concepts of normal curvature, geodesic curvature, principal curvature, gaussian curvature, and mean curvature. The unit normal to a surface,  $\mathbf{N}$ , at a point  $P$ , defines a plane tangent to the surface at  $P$ . Any line through  $P$  in this plane locally determines a curve on the surface, and hence a normal curvature  $\kappa_n$ . The normal curvature will be a maximum in one direction and a minimum in the orthogonal direction.<sup>3</sup> These are called the **principal directions**, and the corresponding normal curvatures  $\kappa_1$  and  $\kappa_2$ , the **principal curvatures**. The quantity  $K = \kappa_1 \kappa_2$  is called the **gaussian curvature**, and the quantity  $H = \frac{1}{2}(\kappa_1 + \kappa_2)$  is called the **mean curvature**. Figure 6 illustrates the connection between gaussian and mean curvature and intuitive ideas about the qualitative shapes of surfaces. A curve through  $P$  that connects two points  $Q$  and  $R$  by the shortest path is called a **geodesic**, and, when it is orthogonally projected onto the tangent plane at  $P$ , it forms (locally) a straight line, or, equivalently, a curve of zero curvature. If *any* curve on the surface through  $P$  is projected onto the tangent plane, the curvature of the resulting planar curve is called the **geodesic curvature**. Geodesic curvature and gaussian curvature are intrinsic properties of surfaces.

The qualitative shape of surfaces is suggested by local contours, but the precise shape is very ambiguous. Perception of figures like the Wire Room (Figure 1) seems to depend on global judgments. Perception of particular elements of the figure is preceded by, or depends upon, perception of the figure as a whole — what the Gestalt psychologists called *Prägnanz*. It is possible to obtain, for example, estimates of surface normals using local information [30]. If the “goodness” of the resulting surface description can be estimated, it should be possible to find a global optimum by variational methods (for example, iterative improvement methods such as steepest descent, or more sophisticated optimization methods such as simulated annealing [31]).

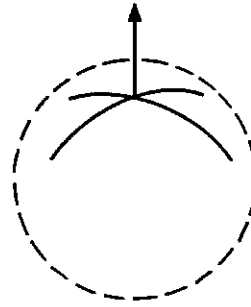
<sup>3</sup>This is not strictly true. The surface may be planar or umbilical at  $P$ , in which case  $\kappa_n$  is uniform.



planar

$$K = \kappa_1 \kappa_2 = 0$$

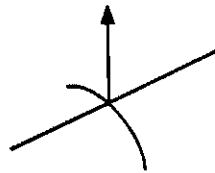
$$H = \frac{\kappa_1 + \kappa_2}{2} = 0$$



umbilical

$$K = \kappa_1 \kappa_2 = \kappa_1^2 = \kappa_2^2$$

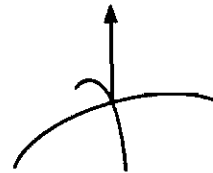
$$H = \kappa_1 = \kappa_2$$



parabolic

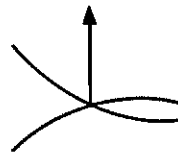
$$K = 0$$

$$H = \frac{\kappa_1}{2}$$



elliptic

$$K > 0$$



hyperbolic

$$K < 0$$

Figure 6: Local Surface Types

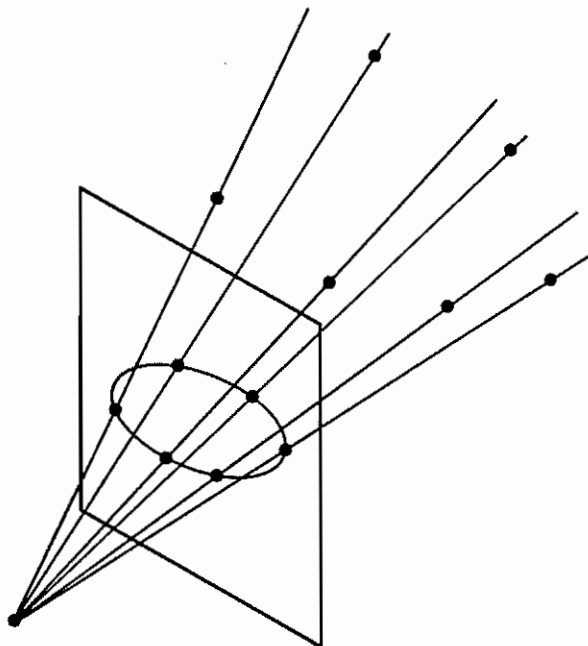


Figure 7: Wire-Bead Backprojection

### 3.2. Generating Hypothetical Descriptions

Even the simplest image represents an infinity of possible 3D scenes. If continuous scene space is quantized appropriately, the discrete space of possible scenes is infinite but denumerable. The class of methods for generating descriptions of these possibilities is **backprojection**. In general, any method that generates three-dimensional descriptions (in terms of distances, orientations, lighting models, reflectance models, etc.) while maintaining consistency with the geometrical and physical constraints of the image, is an instance of backprojection.

Perhaps the easiest way to visualize backprojection is with the “wire-bead” model [32] (Figure 7). Points on the image contour can be backprojected, or placed in 3D space, anywhere along a line connecting the center of projection and the image point. The wire-bead model maintains the most primitive projective constraints, but does not, for example, require connected image contours to backproject to connected 3D contours. A problem with the wire-bead model is that it allows too many degrees of freedom: one for every contour point.

Another form of backprojection is aimed at generating 3D descriptions in terms of different planar orientations (Figure 8). Assuming the image contour is the projection of a more-or-less planar contour in the scene, which is at some indeterminate distance from the observer, planar backprojection generates scale-invariant descriptions of the possible 3D contours. In the simplest case such a system has two degrees of freedom: the coordinates of the unit normal vectors of the planes. Furthermore, if the parameter space is represented as the gaussian sphere (as opposed to gradient space), the space of possibilities is closed — an important property when sampling the space at a finite number of points [11].

Another form of backprojection has been used to find the most orthogonal interpretation of image line segments (see Barnard, [33]). If linear image features can be interpreted as projections of mutually orthogonal lines in 3D space, human observers have a strong tendency to interpret them in this way [34], [35]. The effect is clearly demonstrated in the familiar Ames Room illusion [36]. Line segments can be backprojected to various combinations of orientations (one degree of freedom for each segment), and the combination that

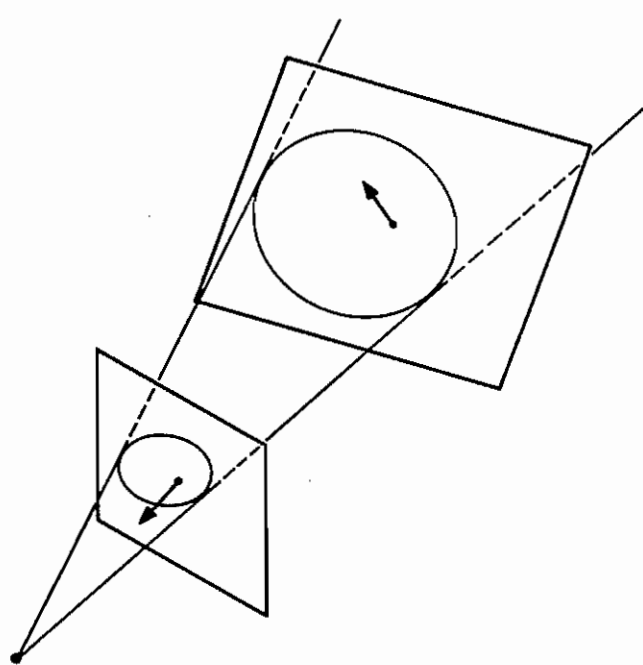


Figure 8: Planar Backprojection

leads to the most orthogonal basis for the vector space of the scene corresponds to the correct interpretation.

It is even possible to extend the concept of backprojection to include illumination and albedo models. The three forms of geometrical backprojection just described generate different shapes from one viewpoint. In addition to varying shape, one could, in principal, vary illumination (for example, by adding or moving point sources), or vary albedo, while satisfying the constraints imposed by the reflectance observed in the image. An image such as the Bumpy Torus (Fig. 2) could be explained in terms of a single-point-source illumination, a uniform albedo, and a smoothly curving surface; or it could be explained in terms of two point sources, implying a shape and/or an albedo that would be very complex. The choice is clear. The problem of using reflectance constraints effectively — connecting the surface shape and albedo to the observed reflectances — is difficult, but there has been promising recent work in this area [37].

In any realistic language, the number of possible encodings of any particular stimulus would likely be enormous. The task of enumerating all of them, while possible in principle, would be hopeless in practice. Information in the primitive encoding, however, may be used to suggest possible forms of final encodings. For example, “T-junctions” suggest occlusion, and sets of lines intersecting at a common point suggest parallelism. In this approach the role of local “cues” is merely to suggest descriptions, but the final interpretation depends only on the form of the descriptions and is not required to account for all the cues.

### 3.3. Levels of Description

Using the formalism of differential geometry, we can, in principle, represent 2D or 3D figures in a precise, well-founded, intuitive way that is independent of the choice of a coordinate system. Section 4 discusses in detail how the simplicity of figures can be estimated from descriptions. The method requires further descriptions at different levels of specificity. We will use the notation developed by Carnap [38].

We assume that a curve or surface has a precise description that captures all aspects of its shape. For example, in the case of smooth, continuous curves, these descriptions consist

of analytic expressions for curvature and torsion. We denote a precise description of this form by  $D^{\text{prec}}$ .

We can convert precise descriptions to approximate individual descriptions (of which there are a finite number) by sampling over the parameter space at a certain precision of measurement. For example, a smooth, continuous curve can be sampled at intervals of arc to yield a sequence of curvature and torsion measurements (to some precision). Denote an individual description by  $D^{\text{ind}}$ , let  $N$  be the number of samples, and let  $K$  be the number of possible distinct measurements. That is, we divide the measurement space (of, for example, curvature and torsion) into  $K$  cells  $Q_j$  ( $j = 1, \dots, K$ ).<sup>4</sup>

Finally, we can convert an individual description to a **statistical** description by counting the number of elements  $N_j$  belonging to each cell. In other words, we can construct a histogram  $D^{\text{st}}$  from  $D^{\text{ind}}$ . The statistical description gives the frequencies of occurrences of the various measurements.

Each level of description is implied by its predecessor:

$$D^{\text{prec}} \Rightarrow D^{\text{ind}} \Rightarrow D^{\text{st}} .$$

Individual descriptions that imply the same statistical description are said to be *statistically equivalent*. A statistical description represents a disjunction of individual descriptions. The simplicity measure that will be described in Section 4 is based on the size of this set.

#### 4. Why are Some Interpretations Preferred?

This approach to figural perception begins with 2D image descriptions that are disordered, or in which the implicit order is hidden, and, through backprojection, proceeds to construct consistent 3D descriptions that may be more ordered. In other words, it works from complex descriptions to simple ones. If 3D descriptions of very simple, highly ordered form are found, they are chosen as the best interpretations. The logical justification for selecting simple descriptions over complex ones is essentially the principle of Occam's Razor.

We can draw a loose analogy with a famous problem of physics. Statistical mechanics provides an explanation, based on probabilistic reasoning, of the behavior of irreversible thermodynamic processes, and, in particular, of the Second Law of Thermodynamics, which states that the entropy of a closed system must increase. In simple terms, closed systems invariably evolve from ordered states to less ordered ones. Boltzman [39] and Gibbs [40] invented the mathematical formalisms of statistical mechanics to account for this. The important insight was to identify entropy, which had hitherto been defined only in terms of macroscopic physical measurements, with probabilistic descriptions of the microscopic states of thermodynamic systems. They were able to show that, because the number of disordered states is vastly greater than the number of ordered ones, the probability of the system moving into a disordered state is extremely high. More recently, Prigogine [41] has further developed the thermodynamic concepts of structure and disorder of complex systems.

In a seminal paper that began the field of information theory, Shannon used the concept of entropy as a measure of information [42]. At first, this seemed to be a completely different concept than thermodynamic entropy, but Brillouin showed that they were closely connected

<sup>4</sup>A *finitization* of this sort happens when a discrete image is created.

and consistent [43], [44]. Jaynes showed that the thermodynamic concept could be derived from Shannon's measure [45], [46].

#### 4.1. A Model of Structure and Information

The property that we use for selecting preferred descriptions is minimum entropy.

Entropy is defined for statistical descriptions, for individual descriptions by implication, and for precise descriptions under some system of finitization. Using the notation developed in Section 3.3, assume we have a statistical description  $D^{\text{st}}$  with cell numbers  $N_1, \dots, N_K$ . The number of statistically equivalent individual descriptions  $D^{\text{ind}}$  with these cell numbers is given by

$$z(D^{\text{st}}) = \frac{N!}{N_1! \cdots N_K!} . \quad (3)$$

The minimum value of  $z$  occurs when all elements belong to the same cell (the homogeneous case):

$$z_{\text{min}} = 1 .$$

The maximum occurs when all cell numbers are as nearly equal as possible (the maximally heterogeneous case). Assuming that  $N$  is divisible by  $K$ :

$$z_{\text{max}} = \frac{N!}{(\frac{N}{K})!^K} .$$

*A system with a statistical description of large  $z$  is more disordered than one with small  $z$ . This is because the statistical description of large  $z$  can be realized in relatively many ways, and it gives us relatively little information about the underlying precise description. On the other hand, if a statistical description has small  $z$ , there are few possible individual descriptions. This observation is the heart of the minimum-entropy principle for figural perception.*

Various sources define entropy in different ways. Shannon, for example, uses the formula:

$$H = - \sum_{j=1}^K p_j \ln p_j , \quad (4)$$

which can be related to  $z$ , the number of statistically equivalent individual descriptions consistent with a  $D^{\text{st}}$ , as follows. We take the probabilities  $p_j$  to be the observed probabilities in a statistical description:

$$p_j = N_j / N .$$

Applying Stirling's formula to (3), we obtain

$$\ln z \approx -N \sum_{j=1}^K p_j \ln p_j , \text{ for large } N .$$

Therefore, from (4),

$$H \approx \frac{\ln z}{N} . \quad (5)$$



The important point is that entropy is always defined as a linear function of the logarithm of  $z$ , even though the details may differ from source to source. The base chosen for the logarithm will affect the units in which entropy is measured, of course, but, since we will only be concerned with comparisons of values, we can use any convenient base and treat entropy as a pure number.

The following definition, given by Carnap, has some useful properties:

$$S(D^{st}) = \ln z - N \ln K . \quad (6)$$

If  $N$  varies but the relative probabilities  $p_i$  do not change, then  $S$  is proportional to  $N$ . Furthermore, if each cell is divided into a fixed number  $q$  of new cells with equal cell numbers  $N_i/q$ , then  $S$  remains unchanged. These properties are computationally attractive because they allow entropies calculated for statistical descriptions with different  $N$  and  $K$  to be compared, which outweighs the minor inconvenience that  $S \leq 0$ .

The concept of entropy is notoriously opaque to intuition. The essential point is that a description will have high entropy when its elements occur with more-or-less the same probability, and it will have low entropy when a few measurements have much higher probabilities than all others. Shannon's measure,  $H$ , can be interpreted as the average amount of information per symbol in a description. An encoding is said to be *efficient* if its symbols occur with equal probability, and therefore carry equal amounts of information, or, equivalently, if the encoded description has maximum entropy. Shannon's original motivation was to discover how to use fixed-bandwidth communication channels most efficiently, and he was therefore led to the concept of entropy as a measure of the efficiency of coding schemes.

The redundancy of a description is defined as:

$$R = 1 - \frac{H}{H_{max}} , \quad (7)$$

or, in terms of Carnap's definition,

$$R = 1 - \frac{S + N \ln K}{S_{max} + N \ln K} . \quad (8)$$

Note that  $R$  is in the interval  $[0, 1]$ , and that  $R = 0$  for an efficient encoding. An encoding with entropy significantly lower than the maximum possible value, however, will contain a degree of redundancy. Finding minimum-entropy interpretations is equivalent to finding maximally redundant ones. Redundancy is thereby *discovered* and can then be exploited to build more concise descriptions.

## 4.2. Some Examples

In this section a few simple examples of the inductive approach will be presented. The minimum entropy criterion will be applied to smooth, continuous, planar (zero-torsion) curves. We will show how various transformations affect the measured disorder of the curves.

Figures 9 to 12 show several curves created with cubic b-splines [47], which, in this case, comprise the precise descriptions of the figures. A cubic b-spline represents a smooth, continuous curve with a finite control polygon, which essentially determines the coefficients

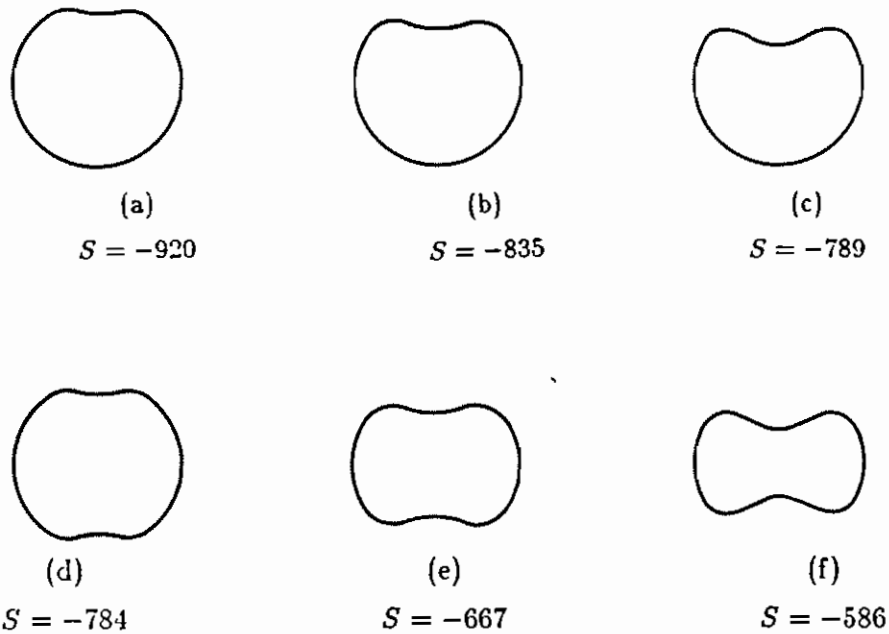


Figure 9: Entropy under Change in Amplitude and Symmetry

of a cubic piecewise polynomial and which can, therefore, be used as an interpolation function [48]. An example of a control polygon is shown in Figure 11k.

To make individual descriptions, the splines are sampled at a predetermined number  $N$  of equally spaced points (500 in all these examples), and curvature is determined analytically from the spline function.<sup>5</sup> A precision of measurement is then chosen (the parameter  $K$ , which was equal to 200 in all the examples).<sup>6</sup>

The first example (Figure 9) shows what happens to the entropy of an initially circular figure as its symmetry is broken, first into a series of three increasingly noncircular figures with one axis of symmetry ((a) through (c)), and then into a series of figures of the same amplitude as the first three, but with two axes of symmetry. Notice that, for a given symmetry, entropy increases monotonically with amplitude. Also, a two-fold symmetric figure has *higher* entropy, and, therefore, is less simple than a one-fold symmetric figure of comparable amplitude (e.g., compare (c) to (f)). This observation shows that Kanizsa's objection to coding theory mentioned in Section 2.2 does not apply to this method. More axes of symmetry do not imply more simplicity. Quite the contrary.

The next example (Figure 10) is another case of symmetry change. All the figures have the same amplitude and only differ by the number of lobes. Entropy monotonically increases with the number of lobes, or, in other words, figures with few axes of symmetry are judged to be simpler than comparable figures with many axes of symmetry. This behavior is quite surprising, because there is no explicit notion of symmetry built into the minimum-entropy

<sup>5</sup>If the precise spline function is not known *a priori*, curvature may be estimated by fitting circles to triplets of adjacent samples of the given figure. In either case, we can also relax the requirement that samples be equally spaced by keeping, as part of the description, the sequence of arc-length segments between unequally spaced samples. Entropy would then be computed using a two-part statistical description: one part for curvatures, and one for arc length.

<sup>6</sup>Before computing individual descriptions for a given set of curves, the interval of admissible measurements must also be fixed. If the bounds are set as tight as possible (i.e., to the actual minimum and maximum of all curvatures of the set of curves), the measurements will be as accurate as possible for a given  $K$ . The same bounds were used in all the examples.

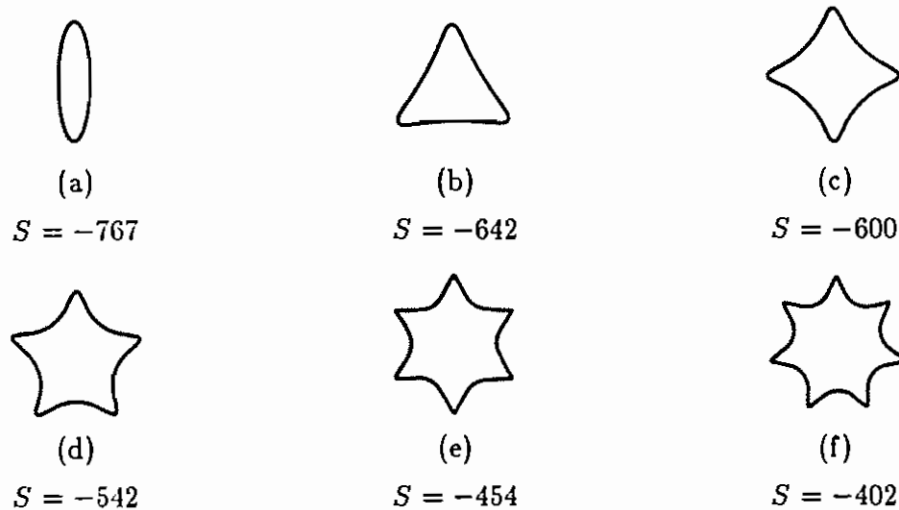


Figure 10: Entropy under Change in Symmetry

model.

If we begin with a highly ordered curve and then introduce random changes, we would expect the curve to become more disordered: entropy should increase. Figure 11 shows that this is indeed the case. The eight vertices of the polygon used to generate an initially circular curve were perturbed by adding zero-mean gaussian noise. A sequence of curves was created by iterating this process. Each curve has undergone twice as many iterations as its predecessor. Entropy increases with the number of iterations — not monotonically, because of the random nature of the experiment (iteration (g) had lower entropy than iteration (f)), but as a statistical trend.

The final example (Figure 12) shows how the minimum-entropy principle can be used to select 3D interpretations. The curve in Figure 9c was rotated in azimuth and elevation and then projected in perspective. The resulting curve, shown in Figure 12a, was backprojected onto several hypothetical planes, which are indicated by tilted circles in the other figures. Just as in the previous examples, individual and statistical descriptions were computed for each of the backprojected figures, and their entropies were determined. As expected, the best interpretation has the lowest entropy, because it corresponds to the interpreted curve that is most regular.

### 4.3. Discussion

The minimum-entropy principle for figural perception expresses a preference for figures that are *simplest* in a certain sense. The measure of simplicity — negative entropy — can be interpreted in several ways, using metaphors of physics, information theory, and inductive reasoning.

Simplicity is the obverse of disorder, which is measured by entropy. Closed physical systems dissolve into disorder; which is to say, they undergo irreversible thermodynamic change. Perceptual systems are not closed, of course. They can freely exchange energy with their supporting systems, and thereby evolve into more ordered states. In a sense,

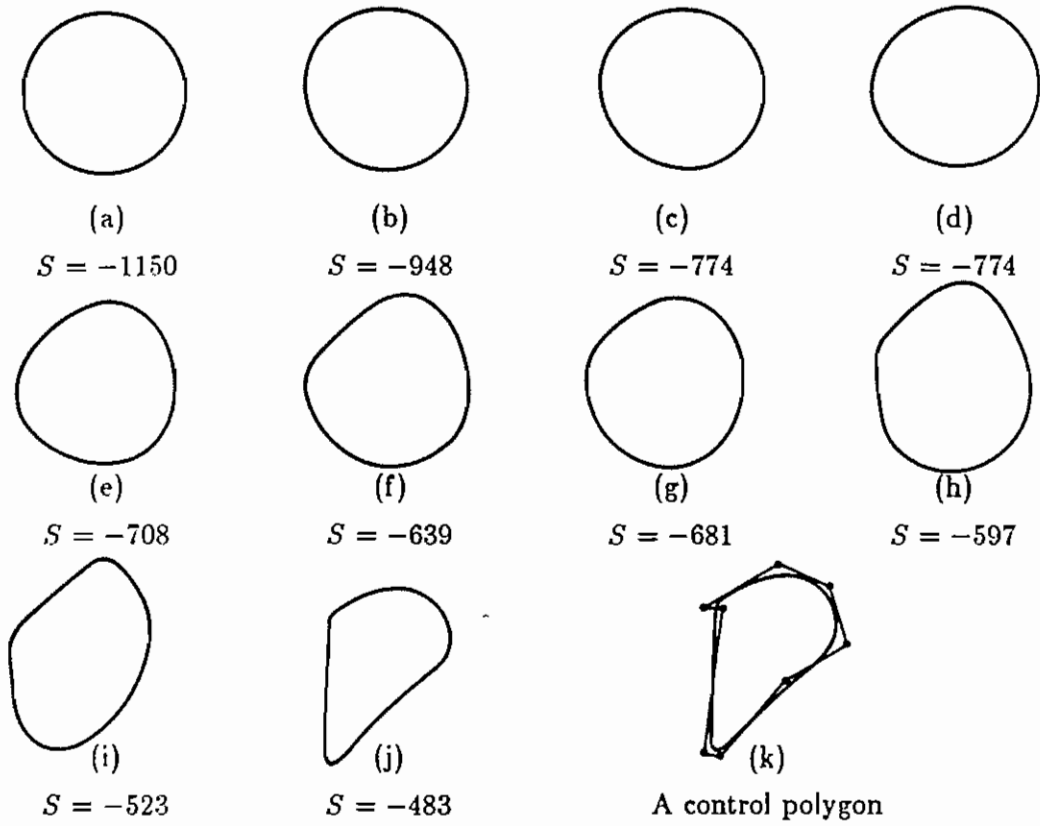


Figure 11: Entropy under Random Perturbation

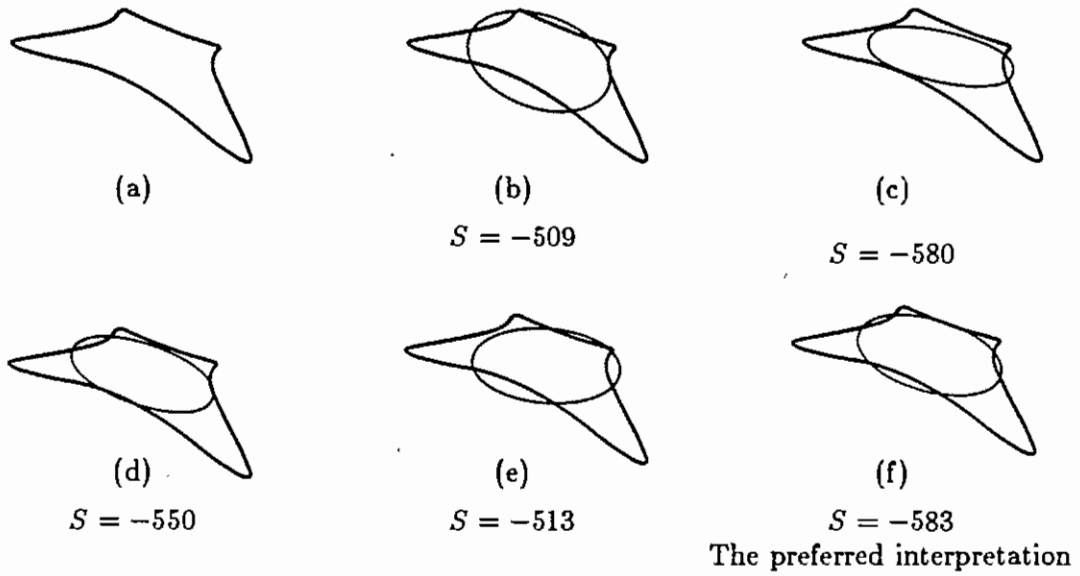


Figure 12: Entropy under Backprojection

the minimum-entropy concept treats perception as the conceptual reversal of physically irreversible processes. Prigogine has developed the concept of entropy exchange to analyze the behavior of open systems [41].

In communication theory, the entropy of a message source is determined by the probabilities of the messages it sends. If there are many more-or-less equally probable messages (high entropy), the receiver is initially in a condition of high uncertainty; if there are relatively few, highly probable messages (low entropy), the receiver has less uncertainty. After receiving the message, the receiver gains an amount of information equal to the uncertainty that is resolved. There are two ways of measuring the amount of information in a message: (a) reduce the message to the shortest possible encoding (i.e., a nonredundant encoding) and then count the number of symbols, or (b) estimate the entropy directly from observed frequencies using Equation (6) and apply Formula (8). The coding theory discussed in Section 2 uses the first method, while the minimum entropy approach uses the second. The advantage to the second method is that it eliminates the need to actually construct a nonredundant encoding — a task that may require considerable cleverness. If we have two individual descriptions with distinct statistical descriptions (but with the same  $N$ ,  $K$ , and bounds), and if one description has lower entropy than the other, then it is more redundant and can, in principle, be encoded with fewer symbols.

The entropic model of complexity, uncertainty, and disorder has profoundly influenced the mathematical foundations of inductive reasoning [49], [38], [50]. The first principle in this foundation has been called the principle of insufficient reason; namely, if there is insufficient reason to believe that several possibilities have different probabilities, one should behave as though they were equally probable. Using entropy as a measure of disorder or as a measure of information follows this principle for the following reason. Given a statistical description, all statistically equivalent individual descriptions are treated as equally probable:

$$P(D_i^{\text{ind}}) = \frac{1}{z(D^{\text{st}})} .$$

If we must choose from a variety of plausible interpretations with different statistical descriptions (e.g., as determined by backprojection), we choose the one leading to the most probable individual descriptions; that is, the one with the lowest entropy.

## 5. Conclusions

The inductive approach suggests a new direction for computational vision. We must face the fact that perception is not veridical and that deductive methods are therefore not appropriate for general-purpose vision. At the same time, approaches that rely on matching specific prior models are unsatisfactory, because they cannot explain the perception of abstract figures of which we have no prior experience, knowledge, or expectation. Recent work toward theories involving a so-called 2.5D sketch (see [51]), when considered as an explanation of figural perception, suffers from the same defect as the deductive approach: there is, in general, insufficient information in a single image to construct iconic, viewer-centered representations of physical surface properties. Relatively direct modes of perception, such as stereo and optic flow, may yield to this approach, but the interpretation of single images will not. Even stereo and optic flow require heuristic assumptions, such as the rigidity constraint, that are closely related to the information-theoretic concept of simplicity.

Induction seems to be a natural paradigm for human intelligence. By observing events, one recognizes correlations, and infers symmetry, causality, family resemblances, and other relationships. To be sure, the inferences may be wrong, but that's too bad. People make mistakes. In fact, one of the weaknesses of deduction is that it does not permit one to draw conclusions that may be in error (assuming the axioms are correct), but that represent the best conclusions under the circumstances.

Only a very small part of a full inductive theory of intelligence is presented in this paper, and several important questions remain to be addressed. For example, one can imagine hierarchies of descriptions, embedded in successively more concise, more global, and more idiosyncratic encoding schemes. To give a trivial example, a curve in the shape of the United States could be encoded as a sequence of arc lengths and curvatures, but it could also be encoded — much more concisely — as a reference to a known shape. How might these hierarchies of descriptions be structured, and how can efficient encoding schemes be learned through experience?

## Bibliography

- [1] Newman, W. M. and Sproull, R. F., **Principles of Interactive Computer Graphics**, McGraw-Hill, New York, 2nd Ed., 1979.
- [2] Mackworth, A. K., How to see a simple world: an exegesis of some computer programs for scene analysis. In *Machine Intelligence 8*, pp. 510-537, N. L. Collins and D. Michie, Eds., American Elsevier, New York, 1977.
- [3] Waltz, D., Understanding line drawings of scenes with shadows, in **The Psychology of Computer Vision**, P. H. Winston, ed., McGraw-Hill, New York, 1975.
- [4] Kanade, T., A theory of origami world, *Artificial Intelligence* **13** (1980), 279-311.
- [5] Rock, I., **The Logic of Perception**, M.I.T. Press, Cambridge, Massachusetts (1983).
- [6] Huffman, D. A., Impossible objects as nonsense sentences, in: B. Meltzler and D. Michie, (Eds.), *Machine Intelligence 6* pp. 295-323, 1971.
- [7] Mackworth, A. K., Interpreting pictures of polyhedral scenes, *Artificial Intelligence* **4** (1973) 121-137.
- [8] Kanade, T., Recovery of the 3-D shape of an object from a single view, *Artificial Intelligence*, **17** (1981) 409-460.
- [9] Kender, J., Shape from texture, Ph.D. Thesis, Carnegie-Mellon University, Computer Science Department, November 1980.
- [10] Ikeuchi, K., Shape from regular patterns (an example of constraint propagation in vision), A.I. Memo 567, M.I.T., Artificial Intelligence Laboratory, March 1980.
- [11] Barnard, S. T., Interpreting perspective images, *Artificial Intelligence* **21** (1983), 435-462.
- [12] Witkin, A. P., Recovering surface shape and orientation from texture, *Artificial Intelligence* **17** (1981) 17-45.
- [13] Brady, M. and A. Yuille, An extremum principle for shape from contour, *Proceedings of IJCAI-83*, 969-972, Karlsruhe, West Germany, August 8-12, 1983.
- [14] Grimson, W.E.L., A computational theory of visual surface interpolation, *Phil. Trans. R. Soc. Lond., B*, **298**, 217-253.
- [15] Grimson, W.E.L., An implementation of a computational theory of visual surface interpolation, *Computer Vision, Graphics, and Image Processing*, **22**, 39-69.
- [16] Grimson, W.E.L., Surface consistency constraints in vision, *Computer Vision, Graphics, and Image Processing*, **24**, 28-51.
- [17] Terzopoulos, D., Multi-Resolution Computation of Visible-Surface Representations, PhD Dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, January 1984.

- [18] Attneave, F., **Applications of Information Theory to Psychology**, Henry Holt and Co., New York (1959).
- [19] Fitts, P. M., Stimulate correlates of visual pattern recognition, *Journal of Experimental Psychology*, 1957, **70**, 21-37.
- [20] Garner, W. R., **Uncertainty and Structure as Psychophysical Concepts**, New York: Wiley, 1962.
- [21] Garner, W. R., **The Processing of Information and Structure**, Lawrence Erlbaum Associates, Potomac, Maryland (1974).
- [22] Hake, H. W., Contributions of Psychology in the Study of Pattern Vision, W.A.D.C. Technical Report 57-62, Dayton, Ohio: Wright Air Development Center, 1957.
- [23] Klemmer, E. T., Perception of linear dot patterns, *Journal of Experimental Psychology*, 1963, **65**, 468-473.
- [24] Quastler, R., **Information Theory in Psychology**, Glencoe, Ill.: Free Press, 1955.
- [25] Buffart, H., Leeuwenberg, E., and Restle, F., Coding theory of visual pattern completion, *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 7, No. 2, April 1981, 241-274.
- [26] Kanizsa, G., **Organization in Vision**, Praeger, New York (1979).
- [27] Borisenko, A. I. and Tarapov, I. E., **Vector and Tensor Analysis**, Dover, New York, 1968.
- [28] Lipschutz, M. M., **Differential Geometry**, McGraw-Hill, New York, 1969.
- [29] Barnard, S. T. and Pentland, A. P., Three-dimensional shape from line drawings, *Proceedings of IJCAI-83*, 1062-1064, Karlsruhe, West Germany, August 8-12, 1983.
- [30] Stevens, K. A., The visual interpretation of surface contours, *Artificial Intelligence* **17** August, 1981, 47-73.
- [31] Kirkpatrick, S., C. D. Gelatt, Jr., and M. P. Vecchi, Optimization by simulated annealing, *Science*, Vol. 220, No. 4598, May 13, 1983, 671-680.
- [32] Barrow, H. G. and Tenenbaum, J. M., Interpreting line drawings as three-dimensional surfaces, *Artificial Intelligence* **17** August, 1981, 75-116.
- [33] Barnard, S. T., Choosing a basis for perceptual space, Technical Note 315, Artificial Intelligence Center, SRI International, Menlo Park, California, January 1984.
- [34] Attneave, F. and Frost, R., The determination of perceived tridimensional orientation by minimum criteria, *Perception & Psychophysics*, vol. 6, no. 6B, December, 1969, 391-396.
- [35] Perkins, D., Visual discrimination between rectangular and nonrectangular parallelepipeds, *Perception & Psychophysics*, vol. 12, no. 5, 1972, 396-400.



- [36] Ittelson, W. H., **The Ames Demonstrations in Perception**, Hafner Publishing Company, New York, 1968.
- [37] Koenderink, J.J. and van Doorn, A.J., Photometric invariants related to solid shape, *Optica Acta*, 1980, vol. 27, no. 7, 981-996.
- [38] Carnap, R., **Two Essays on Entropy**, University of California Press, Berkeley, 1977.
- [39] Boltzman, L., Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. Sitzungsber. d. königl. Akad. d. Wiss., Vienna: 1872. *Reprinted in Boltzman: Wissenschaftliche Abhandlungen*. Leipzig: Barth, 1909, vol. 1.
- [40] Gibbs, J. W., *Elementary Principles in Statistical Mechanics*, Reprinted in *Collected Works and Commentary*, Yale University Press (1936), and by Dover Publications, Inc. (1960).
- [41] Prigogine, I., Unity of physical laws and levels of description, in **Interpretations of Life and Mind**, M. Greene (Ed.), Humanities Press, New York, 1971.
- [42] Shannon, C., *Bell System Tech. J.*, **27**, 379-623. Reprinted in C. E. Shannon and W. Weaver, **The Mathematical Theory of Communication**, University of Illinois Press, Urbana (1949).
- [43] Brillouin, L., Negentropy principle of information, *Journal of Applied Physics*, **24**, N9, 1152-63, September 1953.
- [44] Brillouin, L., **Science and Information Theory**, Academic Press, New York, 1962.
- [45] Jaynes, E. T., Information theory and statistical mechanics, *Physical Review*, **106**, 620, 1957.
- [46] Jaynes, E. T., Information theory and statistical mechanics, *Physical Review*, **108**, 171, 1957.
- [47] Gordon, W. J. and Riesenfeld, R. F., B-spline curves and surfaces, in **Computer Aided Geometric Design**, Robert E. Barnhill and Richard F. Riesenfeld (eds.), Academic Press, 1974, 95-126.
- [48] Barsky, B. A., The Beta-Spline: A Local Representation Based on Shape Parameters and Fundamental Geometric Measures, PhD dissertation, Department of Computer Science, University of Utah, December, 1981.
- [49] Christensen, R., **Foundations of Inductive Reasoning**, published by R. Christensen, Berkeley, 1964.
- [50] Levine, R.D. and Tribus, M. (Eds.), **The Maximum Entropy Formalism**, M.I.T. Press, Cambridge (1979).
- [51] Marr, D., **Vision**, W. H. Freeman, San Francisco (1982).