GRAMMAR FOR VISION

by

LEON GARDNER SHIMAN

A.B., Columbia University
(1958)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June, 1975

Signature redacted

Signature of Author...
Department of Mathematics, May 2, 1975

Signature redacted

Certified by...
Thesis Supervisor

Signature redacted

Accepted by..
Chairman, Interdepartmental Committee

GRAMMAR FOR VISION

by

LEON GARDNER SHIMAN

Submitted to the Department of Mathematics on May 2, 1975 in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Abstract

Our main result is a mathematical model for the study of human visual perception, which we state for perception of icons in the plane. We discuss the relation of this model to judgements about what is being perceived, and to models of visual processes which can be physically tested. We describe the main theoretical concept, stability, both for perception in general and for perception of icons. We illustrate stability by a detailed discussion of a collection of photographs. We develop a mathematical setting in which we can define stability. We then summarize in the stability hypothesis the relation we conjecture exists between the mathematical expression and the underlying intuition of stability. We give a simple computational tool for applying these abstract mathematical structures directly to the study of perception of icons. We point out that our results are consistent with the views of visual artists, they yield a non-trivial classification of types of visual ambiguity, account for a large number of standard "visual illusions", and reveal how to construct many more. We conclude with a new definition of "grammar" which summarizes our underlying methodological considerations.

Thesis Supervisor: Kenneth Hoffman
Title: Professor of Mathematics

Contents

Long before Aristotle's investigation of logical functions of mind, the existence of what we would today call mathematical structures was observed to be fundamental to reason itself. Their existence was confirmed directly by intuition and reflected by grammarians in analyses of language. For millenia we have accepted as appropriate and necessary to the study of language assumed structures of mind. The role of structures of mind in the analytical study of vision has not extended far beyond the definitions of classical geometry.

We present a new technique for investigating the mind-vision domain. This technique yields a formalism which can be mathematically defined. We derive for vision a counterpart to techniques for investigating language known to classical Greece, and still used in studies of language and logic.

We should expect any such technique found applicable to vision to reveal close ties to studies of language with respect to: (i) underlying assumptions about mind, and (ii) methodological problems which arise in using mathematical models to represent judgements of meaning of physical events. Specifically, what are the appropriate structures of mind, known to us only by individual reflection; and what is the appropriate physical setting in which objective correlates to such structures of mind can be found? Although our results are stated in a narrower context, it is our firm belief that these techniques will yield results which bear on long-standing problems central to study of language and psychology.

We have been guided in our investigation by a general principle: clear distinction between mathematics, assumed structures of mind, and physical process. We believe that we can reach a genuine understanding of the mind-vision problem which is beyond cavil, if not dispute, only by constructing a model for vision which is accessible to physical experimentation and thus validation. Our goal has been to find cognitive structures which can be established as necessary to perception and which have an exact mathematical realization which is itself, in turn, embeddable in any model of the physical processes associated with vision.

We present our results in two parts, of which this paper is the first. Its subject is a judgement we call "stability". Somewhat in the way truth is a judgement we render of mathematical statements, stability is a judgement associated with perception of physical objects. We describe an unambiguous procedure whereby judgement of stability associates a mathematical structure with the visual perception of an icon. We call "stability hypothesis" the assertion that a necessary and sufficient condition for perception of an icon is the existence of the stable structure we describe.

Part One can be thought of as "grammar for vision" in that we give a procedure for studying in a mathematical framework meaning-dependent structures which are associated with perception of icons. We will also indicate how, in a natural way, these structures are observer-independent. It is important to note that no physical or real-time considerations are introduced to the grammar, except perhaps to the extent two observers are asked to agree that an object exists.

Part Two, which is to follow shortly, shows the construction of a process model for visual perception in which an icon is a planar light source, and the structures of Part One associated to the icon have a formal realization. We will observe that the model is consistent with full-field, binocular, real-time perception, and yields direct experimental access to the hypothetical structures of Part One.

The focus of our technique is a procedure which associates semantic values to icons. By "icon" we mean a light-modulating surface of finite extent. By "semantic value" we understand an identity assigned to an icon or some part of an icon. We associate semantic values with an icon according to the following rule:

> An identity, which is claimed to correspond to an icon, is admissible only together with specification of the part of the icon with which it is associated. We are concerned only with <u>pairs</u>: values associated with specific parts of the icon.

We make this exact in the following manner:

(i) Using elementary notions of topology, we break a region of the plane into a finite number of pieces, or sub-regions. The region itself we then take to correspond to the icon, the pieces, to parts of the icon which correspond to specific proper values

(ii) Values themselves are represented by symbols.

Their use is subject only to the condition that for a given icon and observer the value of a symbol is always unique.

The stability hypothesis for perception of an icon is contained in the following two conditions:

(i) Each symbol/region pair carries a condition on its associated topological boundary. Those parts of the boundary of a paired region are said to be adjoined to the set if, so far as the corresponding value is concerned, the adjacent region is complete; in other words, if, with respect to the given value, the region is understood to terminate. Those parts of the topological boundary of a region which are adjoined to it will be represented by an arrow pointing toward the interior of the region, and outward, if not adjoined.

(ii) A necessary condition for perception of the icon is that there exists a set of symbol/region pairs which define an exact topological decomposition of the region of the plane taken to represent the icon, and such that the associated boundary arrows always agree; i.e. arrow directions on shared boundaries match.

We call (i) above the <u>adjunction rule</u>, and (ii) the <u>stability condition</u>. The associated collection of symbol/ region pairs we call a <u>stable field</u>.

Stability is a meaning-dependent judgement. Thus, the structures which derive from the stability judgement allow us naturally to define a semantic structure as a particular formal object associated with perception of an icon. We can view a stable field as defining an elementary semantic structure associated to the icon. Any collection of symbol/region pairs assigned to an icon can be studied with respect to the elementary structures they contain.

There are strong empirical grounds for believing that, in general, relative to any stable decomposition of an icon, there exists a full semantic structure associated to it. Evidence compels us to believe that such semantic structures are mathematically interpretable and thus well-defined.

Although it seems to be the case that both mathematically and empirically the associated structures of semantics and the surface features of an icon are distinct, we require the process model of Part Two to make explicit the observation that it is their interaction which constrains the choice of stable decompositions realized for a given icon.

This paper has one main result: we construct a simple mathematical model for perception of icons, without introducing physical assumptions. This provides for the first time a computational tool for studying icons which is consistent with thought and practice in the visual arts. The procedure we employ yields a systematic description of many standard "illusions" considered important to psychology. The form of the result will permit us to bring behavioural considerations into a mathematical model for perception, and systematically to compare perception in biologically distinct systems.

From our main result, we will argue that formal features of the model must represent structures imposed by the mind on perception. Since these structures are defined in a mathematical setting, we can incorporate them for testing in a physical model of visual processes.

The entire theory rests on the judgement of stability. The clarity of that judgement gives us hope that our hypotheses will be verified.

We always identify what falls within our gaze. In familiar surroundings we may know what to expect, so that detailed identification is easy or perfunctory, and may be confirmed by the other senses. We can also observe that an inability to identify what appears before us cannot be ignored.

In this view, perception requires an assignment of identity to the field of vision. We mean by identification an assignment of identity to a specific part of the field of vision. Stability represents the judgement that all parts of the field of vision can be unambiguously identified.

Thus, stability is a natural notion. It refers to an experience basic to study of perception. By nature it is accessible only to individual experience; it depends on a judgement of a state of mind. We can only point toward it; it must be established indirectly, with each reader's independent confirmation.

Our goal is to extract from this notion of stability a mathematical condition on perception. We first restrict our attention to refining and illustrating the relationship of stability to what we have called an unambiguous identification of all parts of the field of vision.

We take identification to be the elementary act of perception*. We think of identification as giving a name to a specific part of the field of vision. This permits us to consider the field of vision broken into a number of pieces, not necessarily distinct, such that each has a corresponding name.

More exactly, we represent identification formally by an ordered pair. The first entry represents an identity; the second entry specifies the region (part) of the field of vision which has been identified. The identity is the attached significance, or semantic value, or value corresponding to the paired region, which by a conscious act of reflection we can ascertain and symbolize, even if not fully communicate or make known to others.

Stability is a condition on a collection of pairs. A

---

* The judgements by which we ascertain compatibility and adjunction are not in this sense acts of perception.

collection judged to be stable covers the field of vision so that all regions have been identified unambiguously. We can then account for observed changes in the field of vision by substituting one collection of identifying pairs for another. All parts of the visual field are accounted for in such a collection. It is important to observe that we cannot tolerate unidentied gaps in the visual field; we must supply a pair whenever we sense that one is missing.

Ambiguity can be given an exact interpretation by introducing a judgement on pairs which we call compatibility. To speak of a pair as an identity assigned to a specific region of the field of vision implies that it makes sense to speak of visualization or visual realization for a pair. If we are given two pairs, then we can specify identity and region for each. There is a judgement which can be rendered of two such pairs. They will be said to be compatible if, together, they can be visually realized without conflict. To be sure, confirmation of this judgement requires that the reader reflect carefully on his own experience.

Given any finite collection of pairs whose regions lie in and together cover the visual field, such that any two pairs can be judged compatible, then we call the collection of pairs consistent, and the corresponding field of vision stable*.

To recapitulate:

- we began from common experience in perception,
- we described the notion of pair,
- we described compatible pairs,
- we showed how a collection of pairs which are mutually compatible and cover the visual field, and thus are said to be consistent, characterize a natural notion of stability.

---

* This is not to be confused with the definition of stable field, which is the subject of section five, although the terms are closely related.

Stability can be studied, albeit imprecisely, by directly applying the judgement of compatibility to perception of icons in the plane. However plausible the foregoing discussion of stability may have been, it is at a level of generality inacessible to direct study.

By restricting ourselves to icons, we can say exactly what we mean by "part of the visual field" for a fixed reference object. If an image of an icon lies within an observer's full field of vision, then by "visual field" we will understand that part of the full field of vision subtended by the icon. This gives a 1-1 correspondence of "parts" of the visual field with regions in the surface of the icon.

The definitions of compatibility and stability, given in section two can now be refined. For an icon, compatibility is defined for identified regions in the icon surface, and stability corresponds to a collection of compatible pairs whose regions fully cover the surface.

Compatibility governs collective identification of regions. For an icon whose regions are easily identified, and can be designated by common names, we can describe and compare individual judgements of compatibility. By breaking the surface of a particular icon into parts which can easily be recognized by name, we can observe directly what is meant by compatibility. The universal applicability of this procedure makes it palin that the phenomenon we have described is quite general.

The fascination of pictures which reveal prominent incompatibilities lies exactly in the fact that they pose a conundrum: what are the obscured relationships between processes of vision and our awareness of what we are actively seeing? Such perverseness in our visual faculty makes itself known in various ways. The variety we find in "visual illusions" is evidence of it. The careful constructions of M. Escher exploit such relations in a methodical, condensed and technically exacting fashion. We are able to account for such perceptual anomalies with the techniques of the stability theory. We have found that their structures are not elementary.

It is very important to bear in mind what we do not say. We do not predict what pairs will be realized in the course of analyzing a given icon. We do not claim that a given observer must, or will most likely, isolate by identification a particular region of an icon, nor having done so, that particular values will be assigned; we could construct icons for which a typical description would be very hard to find. We do not say that stability implies endurance of an image in time. Stability does not imply the absence of transiently-realized pairs. Stability means only that there is a collection of covering pairs which are stably realizable.

Physical conditions which effect perception must be categorically distinguished; they cannot be stated clearly without a model which relates stability to physical processes. We must be careful to restrict our comments to compatible visualization of specific pairs which have common names. But we are not making a statement about names; observations of compatibility are independent of specific names of pairs.

Given a collection of icons, which we might take to be photographs or posters, then we could apply these techniques to each case. Most readers of this text would be able to agree on their observations. We do not discuss *a priori* grounds for the possibility of such agreement. However, it is the case that for a given icon we can say, with some degree of certainty, how a given observer will respond. That will be important to this study in the long run, but it is technically irrelevant to our immediate subject. Our results are independent of cultural bias; but they preserve, formally, effects of prior experience on perception.

The issue of culture-dependence arises in a sense artificially, as a result of our giving an intuitive introduction to stability which is methodologically unprotected. This is corrected in the definition of stable field. Making use of a second judgement, related to but very different from compatibility, we will propose that perception of an icon entails existence of a mathematical structure on the surface of the icon which is called a stable field, that there is a procedure for assigning a stable field to an icon, and that intuitive stability can be realized in the same mathematical structure.

Systematic analysis of icons reveals, from the point of view of stability, distinct types of consistency and inconsistency in their associated formal structures. We postpone discussion of classification; that is more appropriately treated as part of the study of semantic structure, the higher level structures we find naturally associated with stable fields.

We will now illustrate by several cases how the judgement of compatibility is applied to perception of icons. The cases have been chosen for variety, both in the original graphic media and in the associated structures. But their choice was also strongly governed by historical accident and sentiment; several of the cases played important roles in our early investigations. Primarily, however, they were selected because they permit us to make simple and direct observations, sufficiently rich in implication that we can continue to refer to them in subsequent discussions. We have much more to say about these cases, both individually and comparatively, than appears here.

It may be of interest to note that where, as for several of these cases, we have color transparencies which correspond to the black and white prints, that the remarks we make are still valid, and the perceptual phenomena to which they refer even more pronounced.

We can summarize our observations briefly:

- cases A - E reveal prominent incompatibilities,
- case I is a high-quality photograph, rich in detail, without evident incompatibilities,
- case J is not easily stabilized,
- cases K - M are of particular interest for discussion of "visual illusions".

We proceed to a case-by-case analysis. The illustrations we refer to can be found in the Appendix.

In case A, we can easily find pairs which correspond to the names "stretched figure wearing breechcloth and belt" and "head and shoulders of man looking and pointing to the right". We consider them to be incompatible. By contrast, "stretched figure..." and the pair we can identify with the name "huddled group of men" we call compatible.

We can speak of a pair only when the representing region is exactly specifiable. For the pairs we have named in case A, all the associated regions have clearly visible limits. The next case shows us this need not always be true.

In case B, we find the incompatible pairs "(black) holes" and "paint buckets on skylight". In each case both region and identity are easy to ascertain. We can do this despite the absence of a clearly visible bounding line for all of the right hand oval. There is also a pair "bucket sides and handles" which is incompatible with "(black) holes". There are other names, such as "black ball or disk atop column", which seem to be incompatible with "paint buckets"; but the regions are hard to specify, and thus we cannot speak easily of an associated pair.

Case C is clear. "mother and daughter on bicycle" is incompatible with "go out". "mother and daughter on bicycle" is compatible with "g + ut". Note that the region associated with the words and letters of "go out" is exactly the visible part of the surface contained by the (complete or incomplete) letter boundaries.

Case D is harder to discuss. We observe however that "sailing ship" and "gentleman holding loaf" are compatible. Each is incompatible with "visage of woman". This incompatibility is easily verified by assigning the identity of "left eyelid" to the region associated with "white sails".

In case E, "hearing" is associated with a pair whose region is the entire surface of the icon. "ear" names the pair corresponding to the full surface less the part covered by the four letters. We observe that "ear" proves to be incompatible with "hearing". There is a pair "hear" whose region is made up of the region of "ear" and the region of "H". "hear" is compatible with "hearing". We find "ear" to be compatible with "H + ING", whose associated region is the area of the four letters.

Case F presents a subtle incompatibility between "cloud", which includes neither "three small ovals" nor "...mild aromatisch", with "ballooned thought of bee". The region associated with "ballooned thought of bee" is the union of all three of those regions, but the associated identity is incompatible with the identity of "cloud". We observe that there does not seem to be an identity for the region of "three small ovals" which we honestly could represent by the name "three smaller clouds". "three small ovals" is incompatible with "im Garten unter einer schoenen Wolke saeuft Bienlein gluecklich den Honig", whose region we try but find it hard to specify.

In case G, "five biplanes trailing smoke" is incompatible with "long-fingered skeletal hand". "Hitlergruss" is compatible with "long-fingered skeletal hand".

In case H, we observe that the sequence of words "come up to" mark a region in the icon for which we find two distinct incompatible identities which might be named "travel to" and "smoke". They appear subordinate to the incompatible pairs "travel to a mountain river", whose region includes the visible part of the photograph, and "smoke Kools", whose region is marked by the words "come up to Kool".

Case I we find to be without obvious incompatible pairs. It is important to observe that the region belonging to the pair "boiler" excludes entirely the region named "blades of grass", but must include the region named "shadows of grass on boiler". We include this particular case to show how exact must be the treatment of visible detail in specification of a pair. We wish to note explicitly that presence or absence of obviously incompatible pairs is not an _a priori_ criterion of artistic quality or value.

Case J is an icon for which we cannot find obvious pairs which compatibly cover the icon. We cannot find unambiguous identities for regions to represent "smiling model", "sky" or "frame". Assignment of one conflicts with the others. Masking parts of the surface allows partial disambiguation. We observe that "box of Agfa film", "three photographs", the logo and the words can all be unambiguously and compatibly identified. Note that "contorted smiling model, half inside a window" is not an acceptable identification.

For case K, we find the incompatible pairs "facial profile" and "profiled portal"; we also find the compatible pairs "facial profile" and "right hand space", as well as the compatible pairs "profiled portal" and "left hand space". There are thus two disinct ways compatibly to cover the icon.

In case L, we find the pair "left hand facial profile" compatible with "right hand facial profile". But both pairs are incompatible with the pair "baroque 'H'". As in case K, there are (at least) two distinct coverings of the icon with compatible pairs.

In case M, there is no pair "human facial profile", because we cannot specify a corresponding region unambiguously. There is no collection of pairs which cover the icon, either compatibly or incompatibly, which includes the semi-profiles (the upper left and lower right hand sides).

We have thus demonstrated the compatibility judgement for perception of icons in the plane.

We have observed that the compatibility or incompatibility of pairs depends on our ability visually to realize them in an icon. Our purpose was to show that study of compatibility was germane to investigation of perception of icons. In particular, we observed that compatibility characterizes unambiguous perception of an icon.

We refine these techniques with the help of a judgement called adjunction, more elementary than compatibility. Adjunction is defined in a simple mathematical structure. We begin with the definition of what we will call a stable field. The mathematical terminology is standard, and can be found for example in Topology, by J. Munkres*.

By the plane, we mean a Euclidean space of dimension 2, with the usual topology. We consider the collection T ,

---

* Munkres, James R., Topology: a first course, Prentice-Hall, 1975.

consisting of the subsets u with these two properties:

(i) u is an open set of finite extent,

(ii) the topological boundary of u is the union of a finite number of compact analytic arcs. (By compact analytic arc we mean an image in the plane of the closed interval $[0,1]$ under a map which is analytic on a neighborhood of the interval and is 1:1 and has non-vanishing derivative on the open interval $(0,1)$.)

We call T the class of <u>allowable open sets</u>.

It can be proved that the class of sets T is closed under finite union and intersection, and that if $u \in T$ then the interior of the closure of u is again in T.

Let c be an open arc or simple closed curve lying in the boundary of an allowable open set u. We will call any such connected subset c of the boundary of an allowable open set u an <u>arc</u> of the boundary of u. An arc c will be called <u>orientable relative to u</u> if for every point $p \in c$ and for every sufficiently small neighborhood N of p, there is a sub-arc of c which divides N into exactly two non-overlapping open sets, at least one of which lies wholly in u. If c is orientable relative to u, then an <u>orientation of c relative to u</u> is a consistent choice for each point p

in c of exactly one such set, so that each set chosen lies wholly in or out of u. By <u>consistent</u> we mean that if for each of two points in c a suitable neighborhood has been specified and one set chosen, and if the points can be joined by a sub-arc of c which is contained wholly within the two neighborhoods, then the choices of orientation agree wherever the neighborhoods overlap.

<u>First basic definition</u>:

If u is an allowable open set, then an <u>oriented boundary decomposition</u> d for u, consists of:

(i) a finite number of non-overlapping arcs $c_1,\ldots,c_k$ which exhaust the boundary of u, except for a finite number of "missing" points;

(ii) a choice of orientation for each one of the arcs $c_1,\ldots,c_k$;

subject to this (non-triviality) condition:

if c_i and c_j are two oriented arcs of d and if p is a "missing" boundary point such that $c_i \cup c_j \cup \{p\}$ is an arc, then c_i and c_j have opposite orientations.

It can be proved that if u is an allowable open set in T, then there will always exist at least one oriented boundary decomposition d for u.

Second basic definition:

A stable structure is a finite sequence of ordered pairs $(u_1,d_1),(u_2,d_2),\ldots,(u_n,d_n)$ such that

(i) $u_i \in T$ and d_i is an oriented boundary decomposition of u_i ,

(ii) $u_i \cap u_j = \emptyset$, for $i \neq j$,

(iii) if c_i is an arc in the oriented boundary decomposition d_i of u_i , and c_j is an arc in d_j , then $d_i\big|_{c_i \cap c_j} = d_j\big|_{c_i \cap c_j}$; i.e. orientations agree wherever the boundary arcs overlap.

The following definitions give us the basic formal setting in which the stable structures we have defined from elementary properties of the plane can be associated with perception of an icon.

Third basic definition:

A battlefield is a finite set of ordered triples (e,u,d) such that

(i) e is a finite sequence of symbols,

(ii) u is an open set in T ,

(iii) d is an oriented decomposition of the boundary of u .

Fourth basic definition:

A stable field is a finite sequence of triples $(e_1,u_1,d_1),(e_2,u_2,d_2),\ldots,(e_n,u_n,d_n)$ such that

(i) each e_i is a finite sequence of symbols,

(ii) $(u_1,d_1),(u_2,d_2),\ldots,(u_n,d_n)$ is a stable structure.

We will now state a procedure for constructing in the plane a model for a perceived icon I . The model constructed is observer-dependent.

In practical terms, our procedure is equivalent to cutting-up the physical icon with a pair of scissors, breaking the whole into a jigsaw puzzle of parts. Without loss of material, the icon is separated into parts which, when put together again, exactly cover the original object. Each part of the original is found exactly once among the parts.

Fifth basic definition (Euclidean):

Let I be a planar icon of finite extent. We consider a plane which lies on the face of I . We call the surface of I a particular choice of an allowable open set u in the plane which is perceived to correspond to I . We then say v is an allowable set of the surface of I if $v \in T$ and v is a subset of u .

Definition five gives a condition for associating a subset of a Euclidean space with perception of an object. It is important to observe that under the above correspondence it is always possible to find many allowable open sets u which are perceivably indistinguishable representatives of a region in the icon I. Under this correspondence, topological boundaries cannot be seen; their corresponding parts of the icon have physical width zero. Any part of the icon I which can be seen enters this theory as an allowable open set which is paired with an identity.

We can now make precise the concept of symbol/region pair which has been previously discussed. Definition six states the conditions under which an abstract pair a, assigned to an icon I, can be represented by a triple (e,u,d).

<u>Sixth basic definition</u>:

Let a be a symbol/region pair assigned to the icon I. A triple (e,u,d) will be said to <u>represent</u> the pair a if the following conditions are satisfied:

(i) e is a finite sequence of symbols which represents the identity of a,

(ii) $u \in T$ and u is perceivable as the region of a,

(iii) an oriented decomposition d of the boundary of u is determined by the rule: a component c of the boundary of u is oriented inward toward u if the shape of c is determined by the identity corresponding to e .

Condition (iii) is called the <u>adjunction rule</u>. A boundary component c of an allowable open set u is said to be <u>adjoined</u> to the open set u if c is inward-oriented toward u under the adjunction rule.

We wish now to observe that we have spoken of <u>two</u> judgements. One is called compatibility, and defines relations on sets of pairs. The second is called adjunction, and determines orientation of a boundary component of the region of one pair. The compatibility judgement affirms compatible visual realization of two distinct pairs. A boundary component is adjoined to a region if the identity of the region can be said to depend on or be a function of the shape of that component. The second judgement is independent of visual realization of the icon as a whole even though such realization, by providing knowledge of its visible surround, may have helped determine the choice of identity for a region. Adjunction is solely the affirmation of a relation between the shape of a boundary arc and an identity assigned to the region which the arc bounds.

It is important to note that the definition of stable field of an icon (definition eight) does not depend on judgement of compatibility, only on the adjunction rule applied to pairs. We will illustrate adjunction by diagramming stable fields in section seven. The stability hypothesis, which we state below, expresses implicitly an assertion about the relationship of these two judgements. We make further observations on the relationship of adjunction and compatibility in section eight.

Definition seven establishes an observer-dependent, case-specific rule for use of the finite sequences of symbols e . We discuss in detail, in a forthcoming paper, how this condition leads naturally to observer-independent conclusions.

<u>Seventh basic definition</u>:

If I is an icon, then we will say that a collection of pairs $\{a_1, a_2, \ldots, a_n\}$ is assigned to I , and represent the collection by the symbol A , if the following conditions are satisfied:

(i) a_i is a symbol/region pair assigned to I for $i = 1, \ldots, n$,

(ii) there is a collection of triples $\{(e_i,u_i,d_i);$ $i=1,\ldots,n\}$ such that (e_i,u_i,d_i) represents a_i , for $i=1,\ldots,n$,

(iii) $e_i = e_j$ (*i.e.* the finite sequences of symbols are identical) if and only if the corresponding identity of a_i is the same as the identity of a_j .

Eigth basic definition:

If A is a finite collection of pairs $\{a_i; i=1,\ldots,n\}$ assigned to I , then we will say that A defines a stable field for I whenever

(i) the associated collection of triples $\{(e_i,u_i,d_i); i=1,\ldots,n\}$ has the structure of a stable field,

(ii) if u is the surface of I , then $\bigcup_{i=1}^{n} u_i$ is dense in u ; *i.e.* every point of u is either in one of the sets u_i or on the boundary of one of the sets u_i .

We can now state the stability hypothesis:

If I is an icon, then I is stably perceived if and only if there is a collection A of pairs such that A defines a stable field for I .

We now describe a procedure for constructing diagrams of stable structures in T . Diagrams are an important tool for studying the abstract structures associated with perception of icons. A diagram is used in much the same way as a ruler and compass construction in elementary geometry.

It is accepted practice to make sketches which illustrate abstract topological structures. In particular, we can make a sketch of an stable structure in T . This allows us to make a sketch for a stable field of an icon. A diagram is a sketch with special properties. We begin with the definition of a diagram.

Ninth basic definition:

Let I be an icon, and A a stable field for I . A diagram of A is a sketch of the stable structure of A, such that if (u,d) is any pair in the stable structure of A , then

(i) if v is a connected component of u , then v is represented by a visible connected region in the sketch,

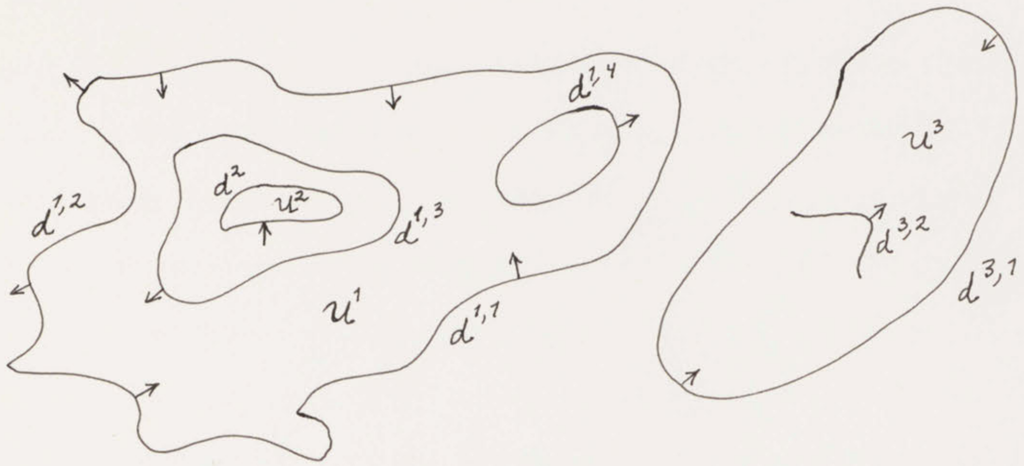
(ii) if c is a connected component of the boundary of u , then c is represented in the sketch by a continuously drawn line which lies adjacent to the region of the sketch which represents u ,

(iii) if c is an arc of the oriented decomposition of the boundary of u , then c is represented by a visible line in the sketch,

(iv) if c is an arc of the oriented decomposition of the boundary of u , then we represent the orientation of c by an arrow; if the boundary of u is adjoined to u along c , then the arrow is drawn to point toward the adjacent interior of the region in the sketch which represents u ; if the boundary of u is not adjoined to u along c , then the arrow is drawn to point away from the interior of the region in the sketch which represents u.

We will call (iv) the <u>arrow rule</u>.

If A is a stable field and (u,d) is in the stable structure of A , then a diagram of (u,d) might typically look like:



where $u = u^1 \cup u^2 \cup u^3$.

Any sketch is perceivable as a planar object of finite extent. Thus <u>it is very important not to confuse the diagram of an icon with the icon whose stable structure it represents</u>. We note explicitly two differences:

(i) visible lines in the sketch have non-zero physical width, but they represent boundaries of open sets in the plane, and as boundaries their width is zero,

(ii) lines in the sketch represent boundary arcs of a stable field; the shape of a drawn line in the sketch need not be and rarely will be like the shape of the arc which it represents.

We wish to make one further observation concerning the relationship of diagrams to the icons whose perceived structure they represent. We do so with the help of a technical fact used also to discuss observer-independence.

Let ϕ be a real analytic homeomorphism of the plane to itself. By that we mean ϕ is a 1:1 real analytic map of the plane into the plane whose inverse is also real analytic and whose differential is everywhere non-zero. We observe that ϕ has the following properties:

(i) if $u \varepsilon T$, then $\phi(u) \varepsilon T$,

(ii) if d is an oriented decomposition of the boundary of u, and c is an arc in d, then $\phi(c)$ is an arc of the boundary of $\phi(u)$, the orientation on c is carried by ϕ onto $\phi(c)$, and ϕ determines uniquely an oriented boundary decomposition for $\phi(u)$,

(iii) if $(u_1,d_1),(u_2,d_2),\ldots,(u_n,d_n)$ is a stable structure in T, then its image in the plane under ϕ is a stable structure in T.

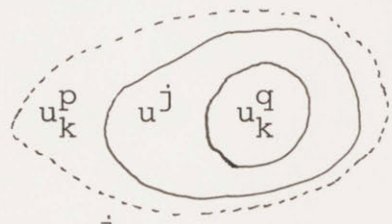
From the previous paragraph we can conclude:

If A is a stable field of an icon I, and B a stable field of an icon J, and there is a real analytic homeomorphism ϕ which maps the stable structure of A into the stable structure of B, then a diagram D of A is also a diagram of B.

For convenience in representing in a sketch an open set u with many disconnected components, we introduce a diagramming convention. The choice of this particular convention will prove convenient in our discussion of semantic structure.

<u>Diagramming convention</u>:

Let A be a stable field and (u,d) a pair in the stable structure of A. If $u = \bigcup_{j=1}^{n} u^j$, u^j connected, then to simplify the drawing of u , we will represent u by the sketch ...  ... if there is a pair (u_k, d_k) in the stable structure of A , so that the diagram of A looks like



for any j for which the diagram of u^j has "holes".

We apply the procedure for diagramming to several cases discussed in section four. The diagrams provide a summary of elementary stability theory.

For cases K and L we have, in each case, two distinct diagrams which correspond to two stable fields. They do not, however, exhaust the list of stable fields which might be associated to the icons. For cases E and H, we sketch one stable field for each icon. For case I, where there are no obvious ambiguities, but many possible stable fields, we only give one stable field.

We observe that for each case the arrows point uniformly outward along the boundary of the icon. This is consistent with the fact that identities for all interior regions are independent of the shape of the icon boundary. We will return to the question of the orientation of the boundary of the surface of the icon.

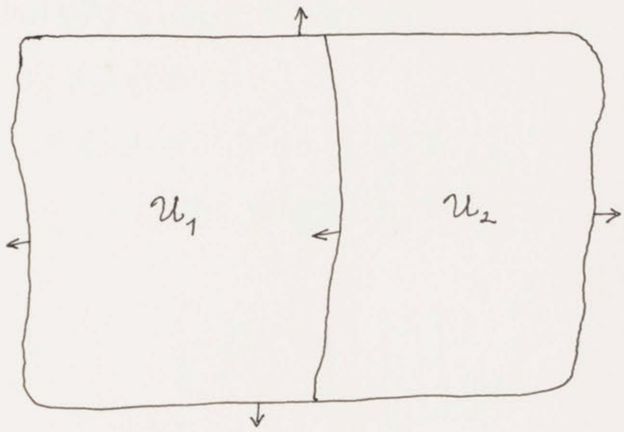
Case K

Let $A_1 = \{a_1, a_2\}$, where

$e_1 = \langle\text{facial,profile}\rangle$,

$e_2 = \langle\text{right,hand,space}\rangle$.

We represent A_1 by the diagram:

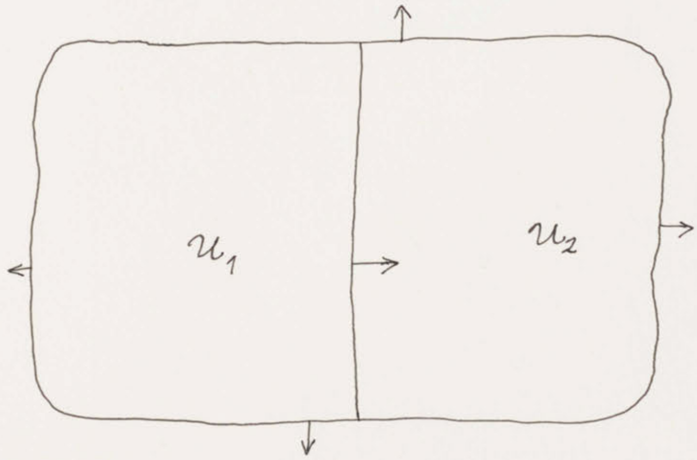


Let $A_2 = \{a_1, a_2\}$, where

$e_1 = \langle\text{right,hand,space}\rangle$,

$e_2 = \langle\text{profiled,portal}\rangle$.

We represent A_2 by the diagram:



Case L

Let $A_1 = \{a_1,a_2,a_3,a_4,a_5,a_6\}$, where

e_1 = <left,hand,facial,profile> ,

e_2 = <space> ,

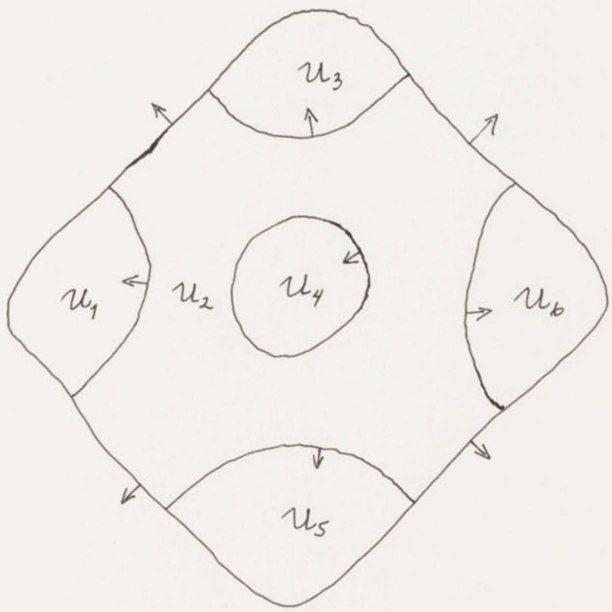
e_3 = <stalactite> ,

e_4 = <floating,object> ,

e_5 = <stalagmite> ,

e_6 = <right,hand,facial,profile> .

We represent A_1 by the diagram:



Case L

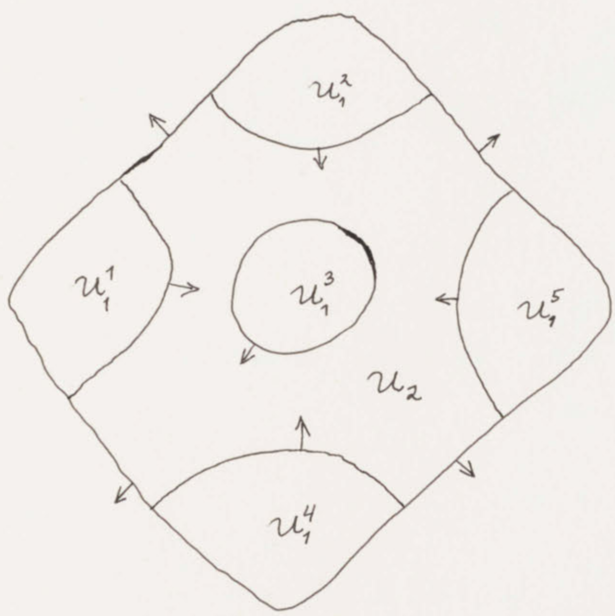
Let $A_2 = \{a_1, a_2\}$, where

e_1 = <space> ,

e_2 = <baroque,H> .

We represent A_2 by the following diagram, where

$u_1 = \bigcup_{j=1}^{5} u_1^j$:



<u>Case E</u>

Let $A = \{a_1, a_2, a_3\}$, where

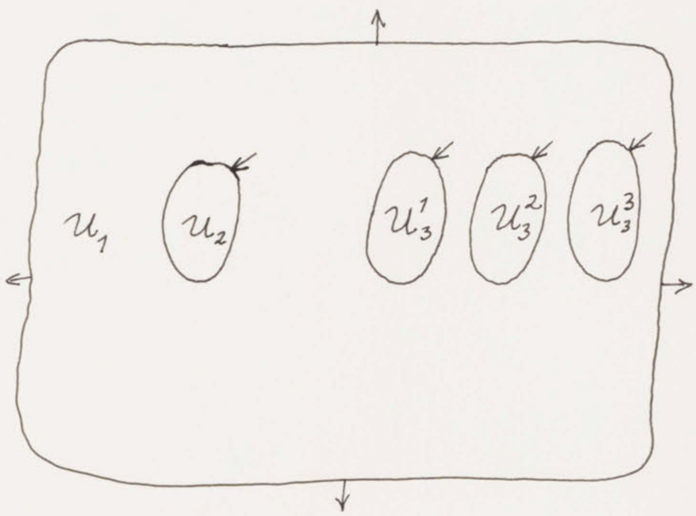
e_1 = <ear,field> ,

e_2 = <H> ,

e_3 = <ING> .

We represent A by the following diagram, where

$u_3 = \bigcup_{j=1}^{3} u_3^j$:



<u>Case H</u>

Let $A = \{a_1, a_2, a_3, a_4, a_5\}$, where

e_1 = <mountain,scene> ,

e_2 = <travel,to> ,

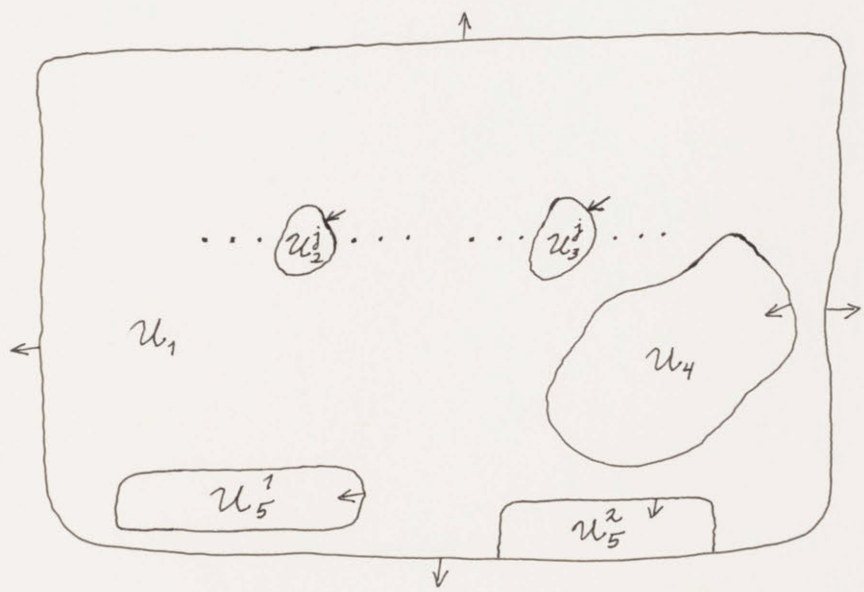
e_3 = <KOOL,brand,cigarettes> ,

e_4 = <packages,of,cigarettes> ,

e_5 = <health,warning> .

We represent A by the following diagram, where

$u_2 = \bigcup_{j=1}^{8} u_2^j$, $u_3 = \bigcup_{j=1}^{3} u_3^j$, and $u_5 = \bigcup_{j=1}^{2} u_5^j$:



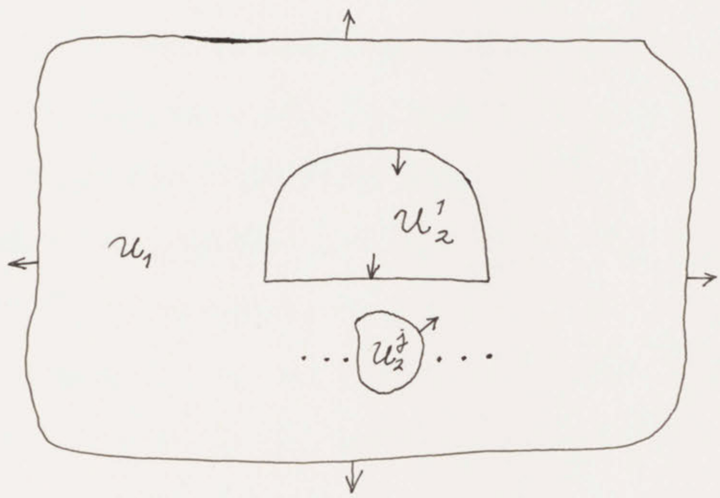
<u>Case I</u>

Let $A = \{a_1, a_2\}$, where

e_1 = <grassy,lot,on,back,street> ,

e_2 = <boiler> .

We represent A by the following diagram, where, for some finite n , $u_2 = \bigcup_{j=1}^{n} u_2^j$:



The model we have given for stable fields of icons can be completed so as to give both a grammar for study of images, and, in a consistent way, a model for visual processes. This can be accomplished in a four step extension of the theory described so far:

(i) We take any finite* list of pairs assigned by a fixed observer to a fixed icon, and by carefully applying the techniques used to model stability, give the entire list an explicit structure. A semantic structure* associated with the icon can be represented explicitly as a structure on a list. We then identify such a list with an image* of the corresponding icon*.

(ii) We get an observer-independent model for perception by formally comparing structures

---

* The assumption of finiteness, and the image/icon distinction will be treated fully in a forthcoming paper.

assigned by two or more observers to the same icon, in such a way that the icon itself can be identified with an equivalence class of images.

(iii) We construct a process model for vision*, which preserves distinct properties of:

(a) abstract semantic structures,

(b) stable structure of an image,

(c) physical features of an icon,

and formally represents their observable relations. The procedure we use respects physical differences between icons, and individual differences in sensory processes of observers.

(iv) We show a natural mathematical completion for the step (iii) model, which incorporates:

(a) full-field vision,

(b) time-dependence,

(c) binocularity,

which, so far as we know, is consistent with fact.

---

* We take as elementary a distinction between models of procedure and process; this is reflected in our usage of "grammar", and will be discussed in a forthcoming paper.

Extensions (i) and (ii) follow directly from the theory we have already presented. Extensions (iii) and (iv) are not *a priori* obvious.

We have called this paper *Grammar for Vision*. Our reasons for having done so are methodological. We summarize those considerations in the following definition.

Tenth basic definition:

By *grammar* we understand a formal model accompanied by a procedure , such that, taken together, the model with the procedure allow us to

(i) relate a mathematical formalism unambiguously to objects of awareness,

(ii) formally preserve a categorical distinction between judgements of states of mind and models of physical processes.

We will identify the stability theory together with extensions (i) and (ii) above with a grammar for vision. We show in Part Two that their completion, summarized in (iii) and (iv) above, gives a model which satisfies our definition of grammar.

The main purpose of this paper was carefully to define and demonstrate the soundness of the notion of stability.

It is central to the entire theory we are proposing. Through this formal characterization of stability we can relate judgements to physical measurements. The theory requires that grammar and process model be used together; the grammar tells us how abstract semantic structures relate to a mathematical description of process. In the formal development of the notion of stability, the two parts of the theory, grammar and process model, are wed. The theory as a whole is only as useful as stability is precise. This careful treatment renders the judgement of stability accessible to the study of vision.

It is important to observe that intuitive stability is as elementary, universally acceptable and unambiguous a judgement as we have of any state of mind. As a direct consequence of our treatment of stability, we can give a mathematical representation to any fact statable in terms stability. Thus any problem of perception which can be so interpreted is directly accessible in the model. From the strength of the judgement itself, we have acquired a useful tool. An extensive analysis of posters suggests that a large class of problems has been made more tractable.

We conclude with three remarks:

First. Our investigations originated with the study of posters. The usefulness of posters to the study of perception derives in part from the following facts:

(i) We are obliged to assign a value* to the whole of the poster, in a reflex-like way.

(ii) Rarely will we have seen all the parts of a poster before, but we are able nonetheless consistently to supply all the parts with identities.

(iii) Although the number of possible parts of posters is unlimited, we are able to find a value for the whole which is consistent with values assigned to the parts.

We inferred from these observations that values are in some way being composed, as a condition on perception of the poster.

Our grammar supplies a formal framework for studying how assigned values are related. We can define (formal) relations on identities by observing how their corresponding sets and oriented boundaries relate, viewed as members of a collection of compatible pairs. This is part of the study of semantic structure.

---

* In the usage of sections one and two.

Second. In the setting provided by the grammar it is easy formally to state a natural classification of types of visual ambiguity. The classification applies as well to icons much more complex in their surface structure than are the icons usually found in collections of "visual illusions". The photomontages of John Heartfield and many of the paintings and drawings of Salvador Dali are good examples. Among the well-known illusions, we can observe directly, in terms of stability, that the "Stimmgabel" (or "trident") can not be stably decomposed in a way which includes either one of the two identities we try to extend from the end regions. The icon of the old-/young-visaged woman with a plume, is of a different type; but it is easy to see that in terms of stability there exist two distinct stable states. Cases K, L, and M were constructed to help discuss their many and varied, well-known, counterparts.

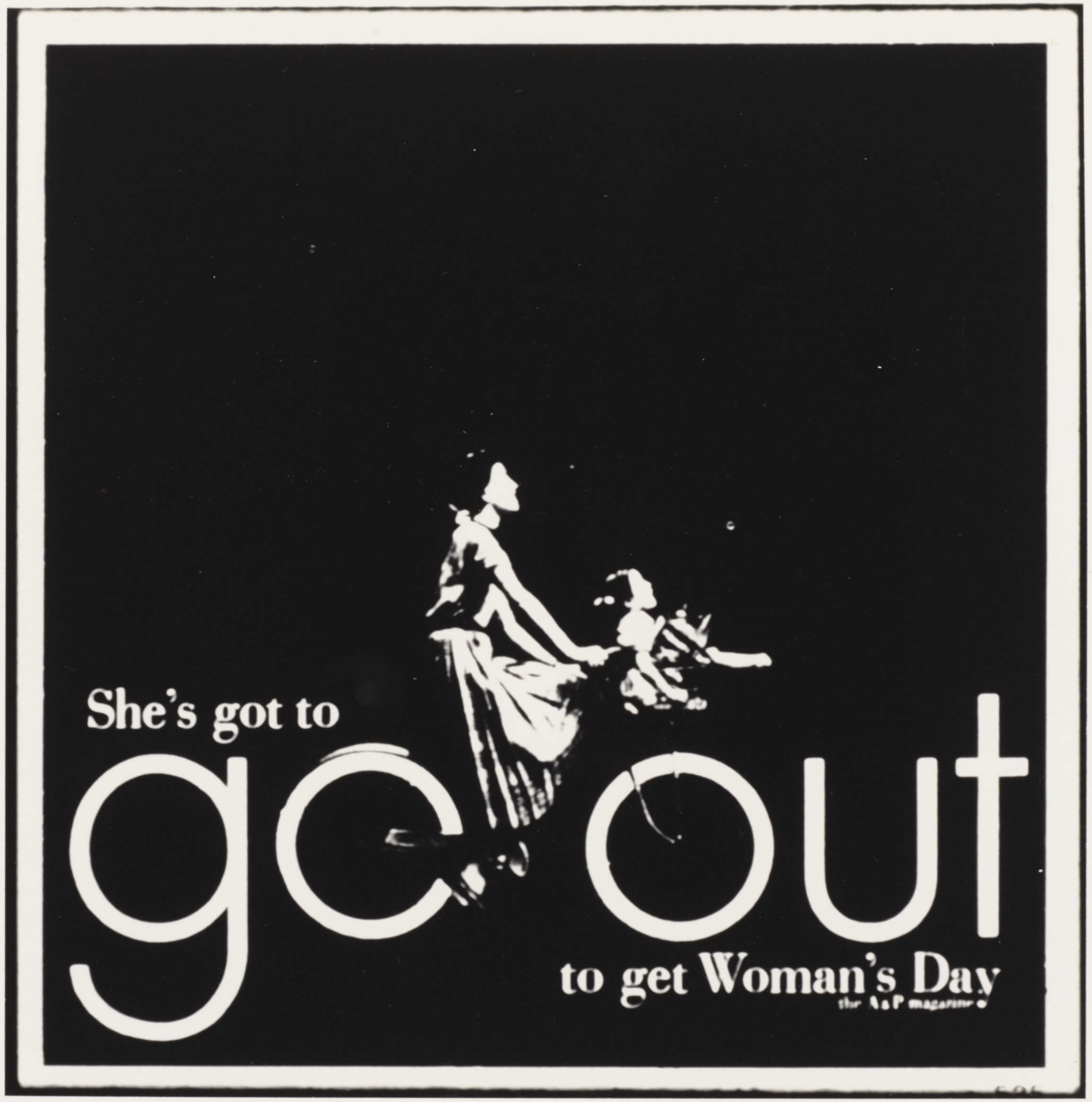
We have not given a detailed analysis of the cases just mentioned, nor of the intersection illusions of Poggendorff, Zoellner, Hering and Wundt; nor of Necker's "cube". The first condition of such analysis is agreement on a procedure for comparing in a formal model of causal relations the structures we assign to particular icons. To do so requires we be able to make statements about the physical structure of the icons, and to be able to compare cases with respect to physical features. We can and will do so in the expanded setting of Part Two.

Third. It is in this way that we find ourselves in need of a model which incorporates process, a model in which abstract structures which represent semantic relations are put in correspondence with an image whose structure satisfies the stability hypothesis, so that the image relates to a physical model of the icon surface. In this way we can discover how modulation of light is related to perception.

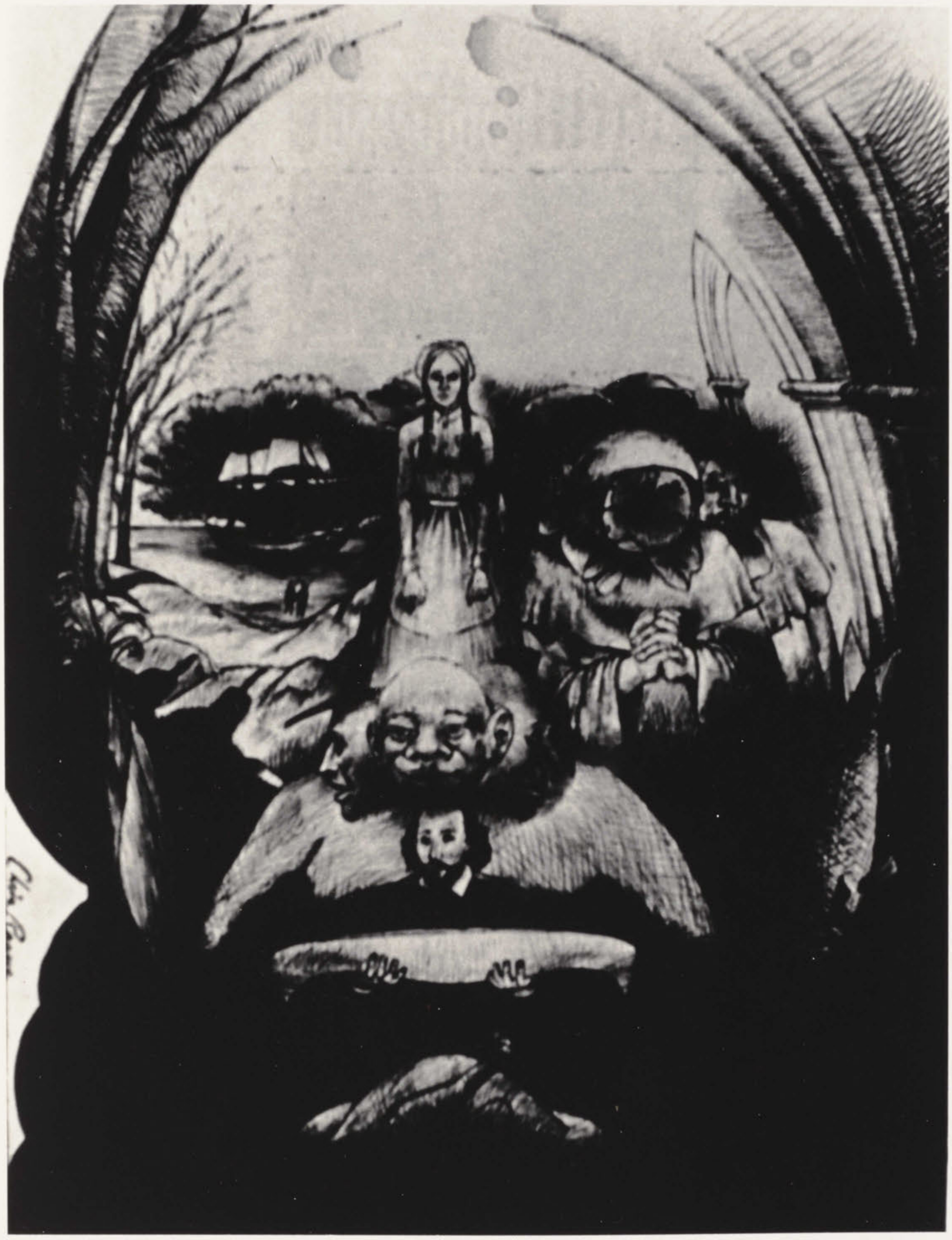
We have learned to think of a stable field as a structure of mind, which has a physiological correlate, and which governs the matching of physical events to an abstract structure which represents significance.


人かたまつて人になる
人おほき人の
中にも人ぞなき
人になれ人
人になせ人
一勇斎國芳戯画

Case B



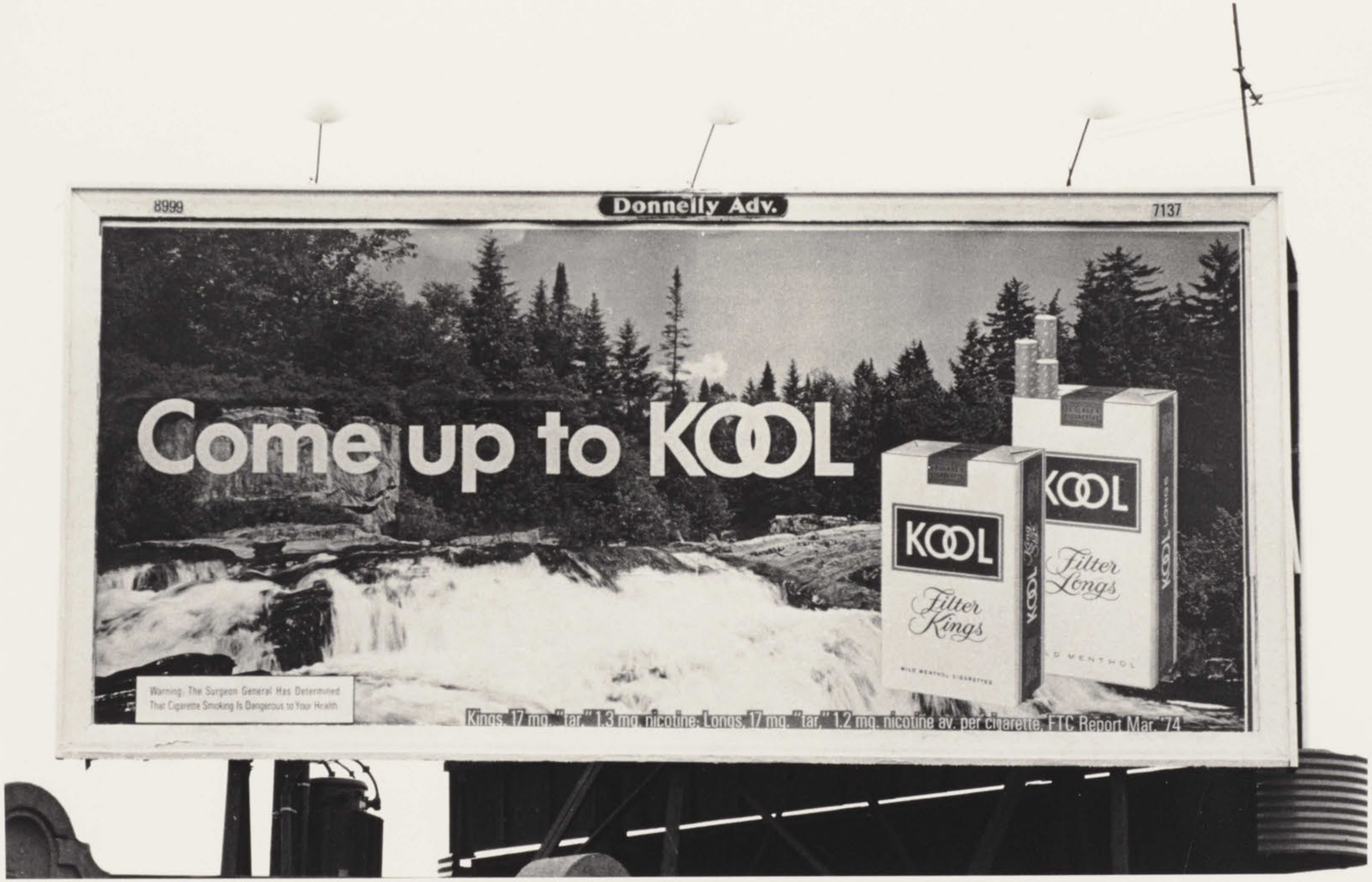
Case C



Case E

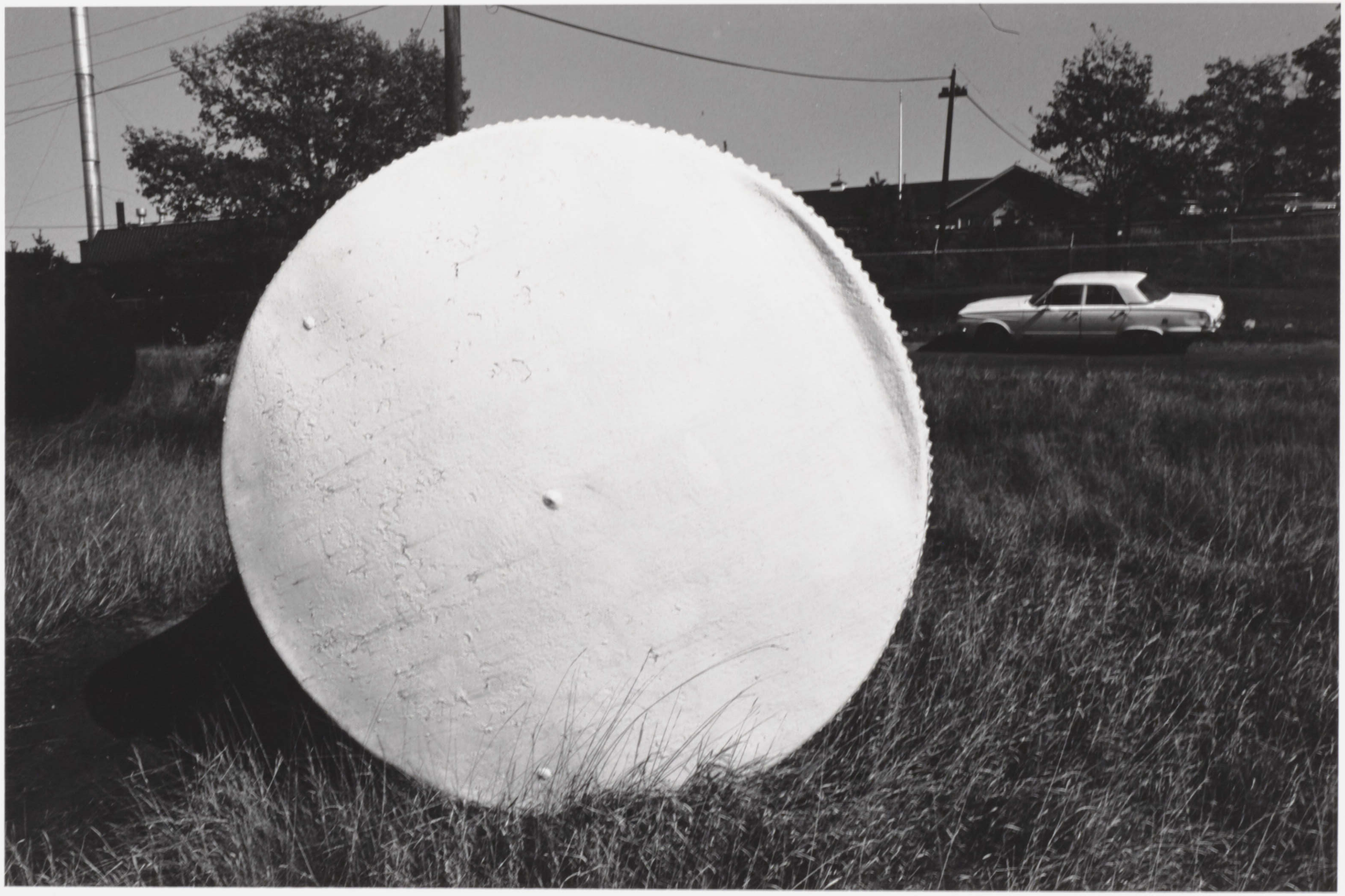

...MILD-
AROMATISCH!
SCHLUSS
SOLO
WILDBLÜTEN-
HONIG


Das ist das Heil, das sie bringen!


8999
Donnelly Adv.
7137
Come up to KOOL
KOOL
Filter Kings
MILD MENTHOL CIGARETTES
KOOL
Filter Longs
Warning: The Surgeon General Has Determined
That Cigarette Smoking Is Dangerous to Your Health
Kings, 17 mg. "tar," 1.3 mg. nicotine; Longs, 17 mg. "tar," 1.2 mg. nicotine av. per cigarette, FTC Report Mar. '74

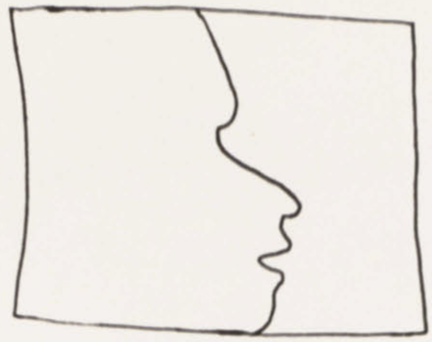
Case H

60




Life Story
Agfacolor
...der Film mit den natürlichen Farben
Agfacolor
SPECIAL
CNS-PAK 126
Agfa
AGFA-GEVAERT

Case J



K

M

L

Credits for photographs

Case A _Men Get Together and Make a Man_, Utagawa Kuniyoshi. From, _The Raymond A. Bidwell Collection of Prints by Utagawa Kuniyoshi_, The Museum of Fine Arts, Springfield, 1968; p. 112.

Case B Photograph by Jonathan W. Green, from his collection.

Case C Gene Federico, for _Woman's Day_ magazine. Reproduced in _Letter and Image_, by Massin; Van Nostrand Reinhold Company, New York, 1970; p. 138.

Case D Reproduced for the author from a source he cannot recall.

Case E Louis Dorfsman. Reproduced in _Letter and Image_, by Massin; Van Nostrand Reinhold Company, New York, 1970; p. 138.

Case F Reproduced from a Kodachrome transparency. Photographed by the author in Berlin (West), Germany; spring, 1972.

Case G Reproduced from _John Heartfield_, by Wieland Herzfelde; VEB Verlag der Kunst, Dresden, D.D.R., 1961; p. 210.

Case H Photographed by the author in Cambridge, Massachusetts; October, 1974.

Case I Photograph by Jonathan W. Green, from his collection.

Case J Reproduced from a Kodachrome transparency. Photographed by the author in Berlin (West), Germany; spring, 1972.

Cases K,L,M Invented and drawn by the author.

Excepting cases B and I , the black and white photographs accompanying this text were made by Susan Taylor. Photographs B and I , were made by Jonathan Green.

Acknowledgements

My particular thanks to the Office of the Provost, M.I.T. for financial support of the photography project during 1971, which made the early stages of this study possible, and to the interest which accompanied that support on the parts of Jerome Wiesner and Walter Rosenblith.

I am indebted to my students, both in the departments of Architecture, M.I.T., and at the Free University in Berlin both for their photography of posters, and for the searching questions which their interest in analysis of graphic images raised in my mind.

This study would not have been possible without the help at different stages in its development of I.M. Singer, Manfred Bierwisch, Henry Millon, Morris Halle and Richard Held.

My debt to my parents, to Susan Taylor and to many friends for personal support is inexpressible.

To my mentor and friend Kenneth Hoffman, I would like to dedicate this work.

Biography

The author was born July 20, 1935, in Washington, D.C. He received the A.B. degree from Columbia University in 1958. He was a student in the Department of Architecture, M.I.T., from 1958 to 1961. He was a graduate student in the Department of Mathematics, M.I.T., from 1967 to 1970, and for the Spring Term, 1975. The author has taught at M.I.T. in the Departments of Architecture and Mathematics, and at the Freie Universitaet zu Berlin (West), in the Fachbereich Philosophie und Sozialwissenschaften.