ELSEVIER

Acta Psychologica 102 (1999) 137–163

acta psychologica

The role of objects in perceptual grouping

Jacob Feldman

Department of Psychology, Center for Cognitive Science, Rutgers University, New Brunswick, NY, 08903, USA

Received 11 May 1998; received in revised form 22 October 1998; accepted 10 November 1998

Abstract

Perceptual organization can be viewed as the selection of the best or "most reasonable" parse of a given scene. However, the principles that determine which interpretation is most reasonable have resisted most attempts to define them formally. This paper summarizes a formal theory of human perceptual organization, called minimal model theory, in which the best interpretation of a given scene is expressed as the formally minimal interpretation in a well-defined space of possible interpretations. We then focus specifically on the role of types of grouping units, in particular the difficult notion of "object". Although grouping is often thought of as the process of dividing the image into objects, most research in perceptual grouping actually focuses on simpler types of units, such as contours and surfaces. Minimal model theory characterizes grouping units at a logical level, demonstrating how formal assumptions about units induce the observer to place a certain preference ranking on interpretations. The theory is then applied to the more subtle problem of objects, culminating in a definition for objects that is formally rigorous but at the same time captures some of the flexibility of human intuitions about objects.

PsycINFO classification: 2323

Keywords: Grouping; Perceptual organization; Prägnanz; Objects

1. The puzzle of grouping

Perceptual grouping is the process by which raw image elements are aggregated into larger and more meaningful collections. Grouping is widely assumed to be early, automatic, and preattentive (Kahneman & Henik, 1981; Prinzmetal & Banks, 1977;

PII: S0001-6918(98)00054-7

Treisman, 1982), though the extent to which grouping can proceed without attention is controversial (Barchilon Ben-Av, Sagi, & Braun, 1992; Rock, Linnett, Grant & Mack, 1992). Grouping and scene organization can impose decisive influences on other low-level processes (e.g., lightness perception, Gilchrist, 1977). Grouping is a necessary precursor to object recognition, because for complexity reasons only well-organized groups, rather than arbitrary subsets of the image, can be compared against as stored object models (Jacobs, 1996). Nevertheless, grouping is certainly one of the least understood problems in vision. This state of affairs reflects the difficulty of precisely formalizing subtle human intuitions about the relative "reasonableness" of candidate groups.

Indeed, notwithstanding the rapidity and effortlessness with which human perceivers perform it, grouping is an extremely difficult problem from a computational point of view. The number of candidate groups in a configuration of n items is equal to the number of subsets and hence is exponential (2^n); the number of partitions (divisions of the n items into disjoint subsets) is a far larger exponential function of n. Many early grouping phenomena, such as the detection of collinearity, are often treated by researchers as local problems in a restricted neighborhood, thus reducing the amount of computation required. However, the more general problem of grouping is well known to involve global effects. Long-distance influences over large areas of the image are common, meaning the fundamental complexity remains extremely high (a fact reflected in the very term "Gestalt", connoting the primacy of the whole). Perhaps the best illustration of the difficulty is the fact that in computational vision, it has become commonplace to require a human user to outline target shapes in images before recognition or motion tracking can commence, because existing grouping algorithms do not provide sufficiently robust or accurate results. The lack of good algorithms in turn reflects the failure of psychologists to propose a theory rigorous and concrete enough to be implemented computationally.

Yet the real theoretical difficulty in grouping stems from the difficulty in clearly defining the computational goal: a rigorous definition of what makes a "good group". Unlike such physically grounded variables as depth, color, and motion, goodness of grouping candidates does not have an objective physical definition. Some ways of combining image elements simply seem more intuitively reasonable than others. The Gestaltists called this elusive quality of perceptual goodness Prägnanz, usually translated as "good form". Fig. 1 provides an example. Of the six line segments in the image, four are grouped into one phenomenal object (a flat square) and two, which are not physically contiguous, into another (a "stick" penetrating the square). The factors that make this particular interpretation the most plausible are subtle and not well understood.

Two general strategies for attacking this problem in the literature can be distinguished. Loosely, some authors seek to explain the *procedure* by which the visual system arrives at its preferred percept – i.e., find a process model – while others attempt to characterize the nature of the preferred percept itself (cf. the distinction between *dynamic* and *static* approaches noted by Van der Helm & Leeuwenberg (1996)). The distinction is related to Marr's well-known (Marr, 1982) division

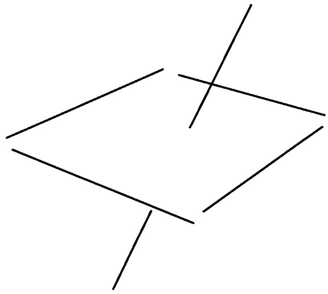


Fig. 1. A line drawing illustrating the role of Prägnanz in selecting the most reasonable interpretation.

between an *algorithmic theory* and a *theory of the computation*, the latter sometimes referred to as a *competence theory* following Chomsky's terminology (see Richards, 1988). As such the two approaches operate at distinct but mutually compatible levels of analysis. The research described in the current paper places the emphasis on the competence theory, on the belief that trying to discover how the visual system computes something – without first defining that thing – amounts to letting the tail wag the dog.

Hence, this paper focuses on an attempt to define in formal terms exactly which interpretation for a given scene is most preferred by human observers, and why. Mathematical details and computational issues in the theory, called minimal model theory, are explained in more detail elsewhere (e.g., see Feldman, 1997c). The emphasis here will be on one particular issue: the role of *grouping units*. What kind of groups – contours, surfaces, objects etc. – are image items aggregated into, and why? In particular I will attempt to shed light on the somewhat amorphous concept of "object", the grouping unit most difficult to define and hence, perhaps, most in need of a rigorous theory.

2. Grouping units

In the common wisdom, perceptual grouping is the process whereby the visual image is decomposed into objects. However, this definition is somewhat at odds with the way perceptual grouping is studied in practice by researchers in the field. More commonly, research has centered around how visual items are organized into striated patterns (Barchilon Ben-Av & Sagi, 1995; Kubovy & Wagemans, 1995; Zucker et al., 1983), contours (Caelli & Umansky, 1976; Feldman, 1996, 1997a; Link & Zucker, 1987; Pizlo et al., 1997; Smits & Vos, 1987), and Moiré patterns (Glass, 1969; Prazdny, 1984; Stevens, 1978). Researchers studying perceptual completion behind a subjective occluder (Kanizsa, 1979; Kellman & Shipley, 1991; Takeichi et al., 1995) or a visible occluder (Buffart, Leuwenberg & Restle, 1981; Sekuler et al., 1994; Van Lier et al., 1995) have usually conceptualized the completed thing as a simple surface (though see Van Lier, 1999). Such an object though is at most a very simple one, consisting of only a single closed region, and almost invariably 2D (though see Tse, 1999). The computational literature has also focused primarily on contours (Guy & Medioni, 1996; Zucker, 1985) and surfaces (Barrow & Tenenbaum, 1981; Binford, 1981). In the human vision literature in general, there is a widespread view promulgated by Gibson (1979) that surfaces rather than objects are the primary unit of visual representation (see He & Nakayama, 1992; Nakayama & Shimojo, 1992).

Objects per se have been little studied in the context of grouping. [1] For the most part this probably stems from the difficulty in precisely defining them. Contours always have a certain well-defined geometrical form: they are 1D space curves, i.e., smooth deformations of the unit line. Similarly, surfaces are always smooth deformations of a neighborhood of the plane. Many objects are simply 3D analogs of contours and surfaces: smoothly bounded regions of 3D space (i.e., "blobs"). In general though objects can be more complex than this, having parts and articulated substructures, and potentially complex spatial relations within them (Fig. 2). Given the difficulty in completely characterizing human grouping preferences even for these geometrically simpler units, grouping researchers have not often approached the more abstract problem of objects directly.

On the other hand, objects have been a central focus in the developmental literature (e.g., Baillargeon, 1994; Spelke, 1990, as well as Xu, 1999). There, interest has centered on defining properties of objects that go beyond strictly visual aspects, such as spatio-temporal cohesion and stability over time. (Indeed a complete definition of objects would certainly combine both visual and non-visual aspects, which is beyond the scope of this paper.) Objects have also been of central interest in the study of attention. They are thought to be the loci of feature binding (Ashby, Prinzmetal, Ivry & Maddox 1996; Treisman & Schmidt, 1982) and of spatial indexes (Pylyshyn, 1989). Moreover, it is known that attention can be moved more easily within than between objects (Baylis & Driver, 1993; Kramer & Jacobson, 1991). The basic question of *which* parts of the visual array can count as objects, though, must be provided by an earlier process, presumably perceptual grouping.

Minimal model theory (henceforth MM theory; Feldman, 1997b,c) has been developed as an approach to the general problem of perceptual organization, not specifically the problem of defining objects, but the theory provides a foundation on which an object definition can be built rather directly. Section 3 gives some background, and Section 4 give a general précis of the theory. Section 5 considers more specifically the role of grouping primitives in the theory, and Section 6 takes up the specific question of objects. In Section 6 a simple formal criterion is proposed that putatively captures the intuitive notion of "object".

[1] Conversely, much research has been devoted to how objects are represented and recognized (e.g., Biederman, 1987; Tarr & Pinker, 1989, as well as Lawson, 1999). But this research has usually assumed that the problem addressed here, namely how objects are separated from other objects and from the background, has already been solved.

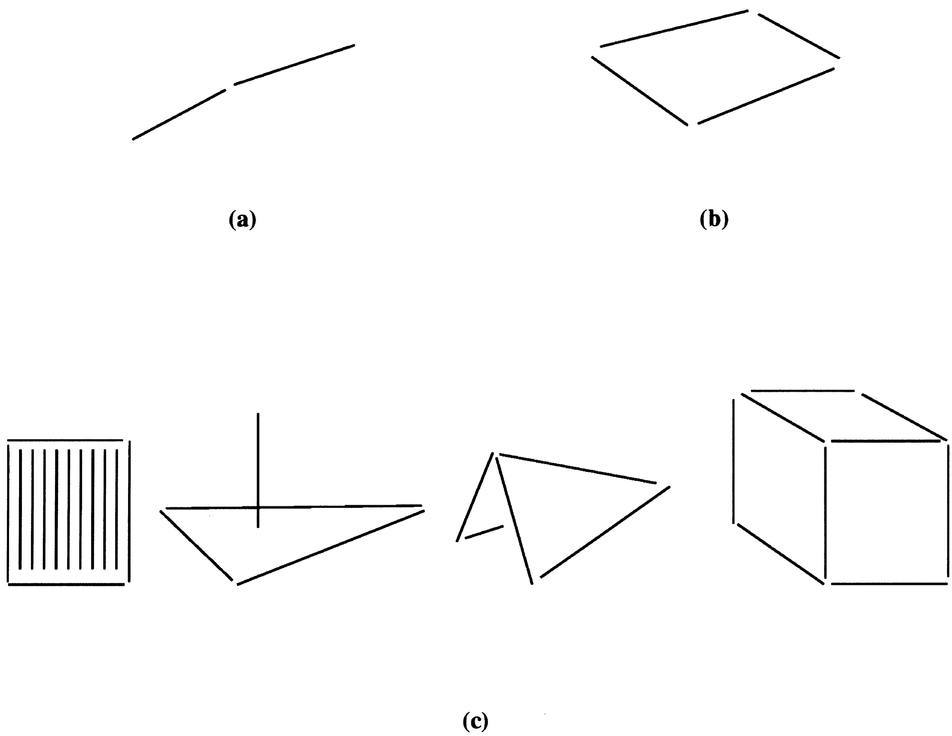


Fig. 2. Line segments grouped into (a) a contour, (b) a surface and (c) complex objects.

3. A "logical" approach to grouping

Our overall goal is to develop a theory describing the rules human observers use to choose the best interpretation of a given scene. Hence, we begin with the problem of defining an "interpretation", a term often used in vision in a loose way but rarely defined carefully.

Reiter and Mackworth (1989) (see also Clowes, 1971) have proposed a definition of an interpretation using ideas from mathematical logic, a field in which the idea of enumerating the alternative interpretations of a fixed set of facts is a central concept. Their definition is quite technical. The discussion here follows their definition only loosely, and is oriented specifically around the idea of choosing grouping units.

Consider an observer presented with a (proximal) scene x, who would like to infer from x a model M of the (distal) world. The observer assumes that the world is made of some type of units P – i.e., that the scene elements proximally observed were actually generated by P's in the world. Further, obviously, the observer assumes that the true world model is observationally consistent with what is observed, i.e., with x. In effect, the observer is making two assumptions about the origins of the observed scene x, which we call *image axioms*:

Axiom 1 (*Observational consistency*). M is consistent with x.

Axiom 2 (*Scene primitives*). M is built from units of type P.

The observer seeks an interpretation that satisfies both of these axioms. In mathematical logic, one speaks of a "model" as a concrete object that satisfies some set of abstract predicates, e.g., the assignment $x = 7$ is a model of the theory "x is prime". Similarly, an interpretation of a scene can be thought of as a model that satisfies the image axioms. Hence, we define:

Definition (*Grouping interpretation*). An *interpretation* is a world model M that satisfies Axioms 1 and 2.

If P is (say) "contours", then this is simply a strict way of saying that the observer seeks a contour-based world model that is consistent with the image. Twenty years ago, a precise statement of the problem in logical terms might have seemed elegant, but would have been of little practical use. Today, however, computational methods for evaluating logical expressions – for example for finding models of logical premises – have reached a high state of development. The field of Logic Programming has progressed from its beginnings in the 1970s to the point where methods for constructing and evaluating logical expressions are widespread, well-understood, and extremely efficient – as exemplified by the widespread use of practical logic programming languages such as Prolog. Hence, the project of finding a model for our image axioms, and thus a model of the scene, is entirely tractable; the main obstacle is in expressing interpretations in a suitable logical language.

Of course, for any given image x there will typically be many models that satisfy the image axioms for a given unit type P. Take for example a dot configuration that we wish to parse into contours (Fig. 3). Any of the depicted assignments of dots to contours, as well a large number of others, is consistent with Axioms 1 and 2.

In addition to a concrete definition for interpretations, then, what is needed is a way to *rank* the interpretations in order of preference. The core of MM theory is this ranking, which takes the technical form of a partial order among models, coupled with a minimum rule for selecting the most preferable model.

4. Minimal model theory

Minimal model theory has three main components: (a) a relational language for describing qualitative interpretations, (b) a ranking among interpretations and (c) a rule for selecting the most-preferred interpretation.

4.1. Parse trees

First we need a language in which to express interpretations. It is widely believed that figural representations are relational (Hock, Tromley & Polmann 1988; Palmer, 1978; Wagemans, Van gool, Swinnen & Van Horebeek, 1993) and hierarchical

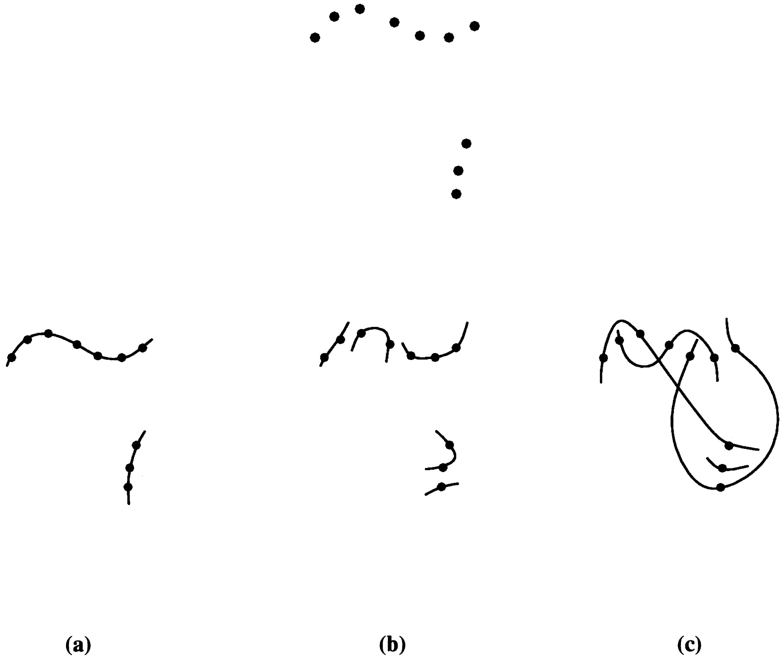


Fig. 3. A dot configuration and several contour models that are consistent with it. Some interpretations (a) are more preferred than others (b, c).

(Palmer, 1977; Van der Helm & Leeuwenberg, 1991). In logical terms a relation such as *collinear* between two line segments x and y can be expressed as a predicate with two arguments, e.g.,

`collinear(x,y)`.

In many cases, the arguments to such a predicate will themselves be complex objects, expressed in the same relational language. For example, the two collinear line segments x and y described by the above expression form a "virtual" line segment (see Fig. 4). This line segment in turn may be perpendicular to a third line segment z. This state of affairs is most conveniently expressed by a tree (see figure), which we call a *parse tree*.

In what follows it is assumed that a complete *qualitative* representation of an observed configuration can be captured in a suitable parse tree – i.e., that qualitative relations among image items are paramount. This emphasis on qualitative spatial relations makes this interpretation language unsuitable for applications requiring metric information, such as face recognition, but it makes it suitable for representing organizational properties such as how the scene should be grouped. [2] The inter-

[2] Of course, metric and qualitative representations are intimately related. For example, qualitative relations such as non-accidental properties (see discussion below), can take the form of special values along measurable parameters (e.g., collinearity, right angles, etc.).

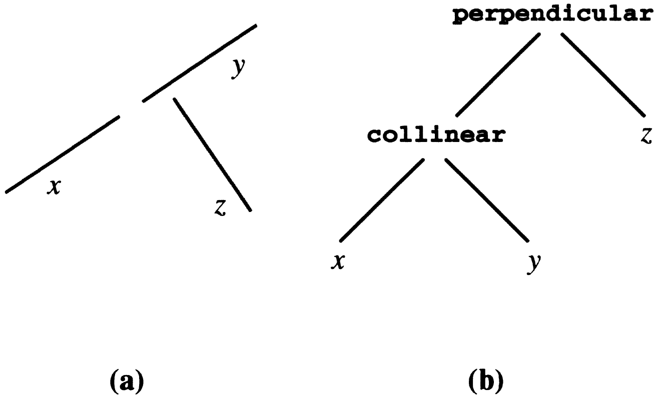


Fig. 4. (a) A line drawing and (b) a parse tree describing it.

pretation problem now reduces to the problem of selecting the most suitable parse tree for a given configuration.

4.2. The partial order among parses

Next we seek to impose an ordering among parse trees, towards the end of formulating a rule for selecting a most-preferred interpretation.

Parse trees are discrete, qualitative objects. The natural formal structure for ranking such objects is a *partial order*. Formally, a partial order (denoted by $\leqslant$) is a reflexive, antisymmetric, transitive relation (see Davey & Priestley, 1990 for an introduction). Informally, a partial order is an ordering in which not all items are necessarily ranked with respect to all other items. (This is in contrast to a *total order*, in which all items are ordered with respect to all others; an example is the integers under their usual ordering.)

An example of a partially ordered relation is the relation *ancestor*: *father* is an ancestor of *child*, and *mother* is an ancestor of *child*, but *mother* and *father* are not ordered with respect to each other (i.e., *mother* is not ancestor of *father* and *father* is not ancestor of *mother*). It is convenient to display orders in a diagram. Fig. 5 gives diagrams (known as Hasse diagrams) of several typical total and partial orders. The crucial idea is to generalize away from simple quantitative comparison (the way we usually use the symbol $\leqslant$) and think instead in terms of abstract rankings.

An important example of a partial order, which is used directly in the theory below, is the relation *subset* (denoted $\subseteq$). For example the four relational facts

$$\{a\} \subseteq \{a,b\}$$
$$\{b\} \subseteq \{a,b\}$$
$$\{\} \subseteq \{a\}$$
$$\{\} \subseteq \{b\}$$

define a partial order that is depicted in Fig. 5c, taking $x \leqslant y$ if and only if $y \subseteq x$.

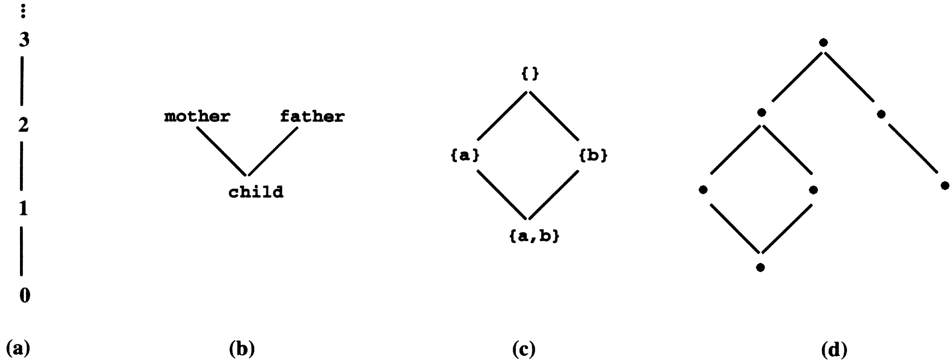


Fig. 5. (a) A total order (the integers) and several examples of partial orders: (b) ancestor, (c) subset, (d) an arbitrary partially ordered relation. For any two items x and y, if x is "hanging" from y in the diagram (even with intervening nodes) then $x \leqslant y$.

In a parse tree, each leaf (bottom node) is an image element (e.g., a dot or line). All other nodes are *models* which act as a head term over several arguments (its children). (Most predicates in this paper have two arguments, but the theory can be generalized easily to any number.) In the example above, the models (`collinear` and `perpendicular`) were "regularities". A regularity is a qualitatively "special" configuration, such as collinear, perpendicular, coterminous, parallel, etc. (cf. non-accidental properties, and see below for discussion.) More generally, models are *sets* of regularities, because more than one regularity can apply at a time. Hence, some models are subsets of other models. Larger models denote configurations that are *more regular* than small models. For example, the model $\{\texttt{coterminous}\}$ is a subset of the model $\{\texttt{collinear}, \texttt{coterminous}\}$ (Fig. 6) and hence describes a less regular configuration.

Hence, models can be ordered by the subset ordering induced on the sets of regularities the models contain. Now, parse trees are made up of multiple models at different levels of the tree. We create a partial order on parse trees recursively, by the following definition.

Definition (*Partial order* $[\leqslant]$ *among parse trees*). For any two trees T_1 and T_2 such that

$$T_1 = \begin{array}{c} M_1 \\ \diagup \quad \diagdown \\ T_{1a} \qquad T_{1b} \end{array}, \qquad T_2 = \begin{array}{c} M_2 \\ \diagup \quad \diagdown \\ T_{2a} \qquad T_{2b} \end{array},$$

we say $T_1 \leqslant T_2$ if and only if

(a) $M_1 \supseteq M_2$, and

(b) $T_{1a} \leqslant T_{2a}$ and $T_{1b} \leqslant T_{2b}$.

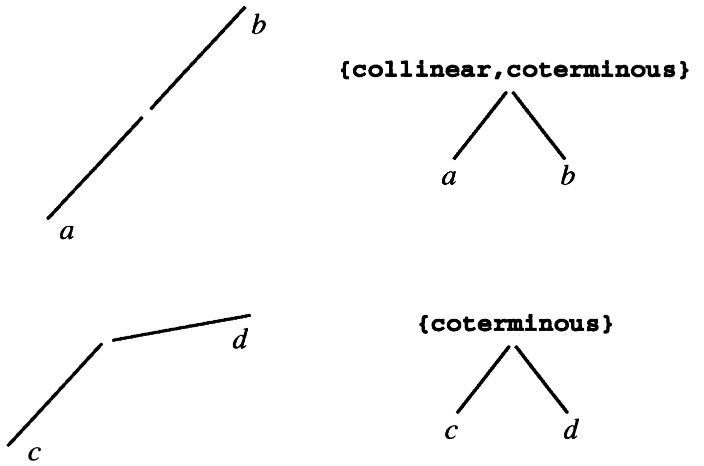


Fig. 6. Two configurations and their corresponding parse trees. The head term of the lower tree is a subset of that of the upper tree, capturing how the lower configuration is a less regular configuration than the upper configuration.

That is, one tree is earlier in the partial order than another if all of its corresponding nodes are more regular (or at least as regular). For example, if two trees are the same except that one's head term is a superset of the other's, then the former is earlier in the partial order than the latter; likewise if the head terms are the same but some internal node is a superset of the corresponding superset on the other tree. Earlier trees are depicted as lower in the corresponding diagram, and from here on we will use "lower" to mean earlier and "higher" to mean later.

The set of all parse trees that are possible (for a given type of configuration, with a given set of regularity types), ordered by $\leqslant$, is called the *interpretation space*. The interpretation space includes all qualitatively distinct interpretations that can be placed on configurations; this is the set of conclusions from which the observer must choose. The interpretation space has a number of interesting formal properties, some of which are discussed and proved in Feldman (1997c). For our purposes here the important point is that the partial order can be shown to be a *preference ranking* among interpretations – lower interpretations should be preferred to higher ones. This crucial fact, which will be discussed at length in the next section, suggests the use of a minimum rule to select the most preferred interpretation.

4.3. The minimum rule

By definition, interpretations that are lower in the partial order are *more regular* than higher ones. For example, say the head term of one contains all the regularities contained in the head term of the other, plus one. In this case the former tree expresses a configuration that is one "notch" more regular than the latter, but is otherwise the same.

A priori, however, it is not obvious why more regular interpretations should be preferred. The world is not always a simple place – at least, it is not always as simple as possible. This point has long troubled philosophers of induction (Quine, 1965; Sober, 1975) seeking to justify Occam's razor. In perception, the role of regularity in perceptual organization is subtle and complex (see Kanizsa, 1979). Although it is not entirely clear why, it is often observed that – all else being equal – a more regular interpretation is preferred to a less regular one. This is precisely the situation represented by two minimally distinct parse trees connected by a line (in graph theory, called an "edge") in the interpretation space. Hence, in the context of a formal theory it would be desirable to explain why two such trees should be preference-ordered – i.e., why the $\leqslant$-ordering should predict intuitive preference.

The significance of the $\leqslant$-ordering depends on the meaning of the regularity terms. Normally, each regularity denotes some kind of non-accidental property (Binford, 1981; Lowe, 1987; see also Wagemans, 1992) – a configuration, like collinearity, cotermination, parallelism, perpendicularity, etc., that is unlikely to be satisfied by accident. In Barlow's elegant phrase (Barlow, 1994), such configurations are *suspicious coincidences*: they rarely occur randomly, and are significant when they do occur (see also Feldman, 1997c; Jepson & Richards, 1992; Richards, Jepson & Feldman, 1996 for discussion). Moreover, though regularities rarely occur by accident, they *typically* occur when some particular world structure occurs – e.g., collinearity among dots is atypical among random dots but typical among dots along a contour (Feldman, 1996). We say the regularity is *non-generic* when a particular world model M is absent, but *generic* when it is present. Hence, when a particular regularity is satisfied by the image, the reasonable inference is that the correct model of the world should include it.

If regularity terms are chosen to have this "suspicious" property, then the $\leqslant$-ordering becomes a preference ordering: whenever a lower model is satisfied it should be preferred. This directly suggests a rule for selecting the most preferred interpretation: the optimal interpretation is the *lowest* (minimal) interpretation that the image configuration satisfies.

Definition (*Minimum rule*). Given a configuration x, among all parse trees in the interpretation space that x satisfies, choose the one that is minimal in the partial order. (See Feldman, 1997d.)

This winning interpretation will be referred to as the *minimal interpretation* (or *minimal parse* or *minimal model*). One can assign to each parse tree T a numeric score, denoted $depth(T)$, measuring "how regular" the interpretation is. [3] For a simple model M, $depth(M)$ is simply $|M|$ (the number of regularities contained in M). For a tree, the depth is the recursive sum of the depth of its head term plus the depth of all of its subtrees. The depth of the generic model is always zero ($depth(\{\}) = 0$).

[3] In other treatments this number is called the "codimension", a term emphasizing a geometric interpretation not important in this paper.

Eliding a few formal details, finding the minimal interpretation is the same as finding the maximum-depth interpretation, and the two terms will be used interchangeably from here on.

Because of the way it is defined, the minimal interpretation is the most preferred grouping interpretation available to the observer. More specifically it is the "least coincidental" interpretation: the interpretation in which the fewest possible number of suspicious coincidences are left unexplained (cf. Rock's "coincidence explanation principle" (Rock, 1983)). Putting it the opposite way, it is the interpretation that *explains* the largest possible number of coincidences. Collinear lines are explained as contours (under which they are generic, or expected); coterminous segments are explained as being joined in the world (in which case one would expect to see them coterminate in the image); and so forth.

4.4. Discussion of the theory

The minimal interpretation has a number of attractive properties. Most importantly, it appears to be descriptively correct as a theory of human perceptual grouping. Several examples are given in Fig. 7, using scenes made of dots, lines, and edge fragments. In each case the minimal parse tree captures the grouping intuition correctly. The winning tree also in effect provides figure-ground segregation; the most salient or structured part of the image, i.e., the figure or "object" corresponds to the maximum-depth subtree in the winning tree. Feldman (1997b) gives examples involving images with hundreds of items, where the procedure correctly finds the psychologically most salient curve. The performance of the algorithm here is comparable to other algorithms designed in an ad hoc manner solely for this problem (e.g., Ullman & Shashua, 1988), while the complexity ($O(n^2)$ for n items) is comparable or better.

Representing interpretations as trees is advantageous in that the tree vividly and directly portrays the qualitative or categorical spatial relations perceived in the figure – bringing the system a step closer to a phenomenal description such as "a stick penetrating a surface". By contrast the usual output of computational grouping algorithms is a processed image which still requires some further (unspecified) processing before it can be described or comprehended.

One advantage of the "logical" characterization of preferences is that grouping of different types of image elements (e.g., dots, lines, edge fragments) receive a relatively uniform treatment. Collinearity, for example, has an intrinsically different definition over dots (where it requires three arguments) than over edge fragments (where it only requires two). Yet in MM theory every collinear configuration (i.e., subtree with the predicate `collinear` in the head position), regardless of the nature of its leaves (i.e., type of image elements), is treated equivalently from the point of view of higher nodes in the tree. That is, one does not need to postulate an entirely different process for handling dots from the one that handles lines, etc., as is commonplace in the computational literature.

One of the main advantages of the theory is that *local* and *global* configural goodness are put into a uniform framework. Imagine that an entire scene is described

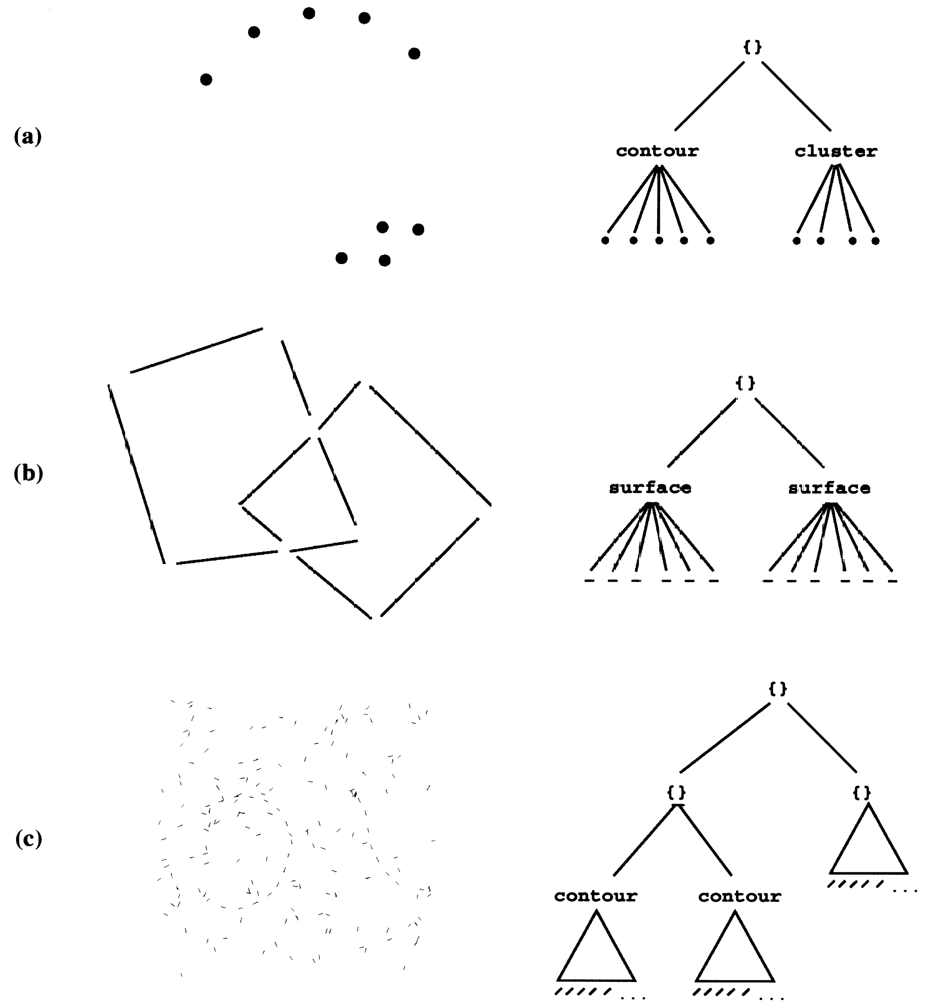


Fig. 7. Examples of configurations (left) and their minimal models (with syntax simplified for ease of presentation): (a) dot configuration, (b) line drawing, (c) edge image. Example (c) shows how the figure or "object" in a field of noise corresponds to a high-depth subtree in the minimal parse.

by a tree T. A subtree low down near the leaves might describe a relation (e.g., collinearity) between two nearby image elements – a local configuration. But at or near the top of the tree, predicates (e.g., symmetry) can refer to large portions of the field, a highly global representation. This is essentially the point famously made by Marr and Nishihara (1978), that configural relations occur at all spatial scales. MM theory gives away of representing these relations that is equally applicable across scales. This may seem obvious, but in the literature there is no similarity or connection in the way (for example) contours are represented and the way global symmetries are represented; the MM argument is that both of these types of regularities, when observed, tend to induce perceptual grouping, and it makes sense to make explicit the structural analogy between them.

The way that MM theory represents proximity is unusual, and requires some explanation. Fig. 7a contains an example of a dot cluster grouped by proximity; the regularity predicate invoked here is usually denoted `coincident`. Just as the predicate `collinear` involves the coincidence of two orientations – but only approximately, as slight angular deviations from perfect collinearity are acceptable – the predicate `coincident` involves the coincidence of two *positions* – but again, only approximately. Just as `collinear` really means "collinear or almost collinear", `coincident` means "coincident or almost coincident", and hence might be better rendered in English as "proximate" (of course, the character string used to denote the predicate is not as important as the predicate's definition!). In either case, the reasoning is the same. If two oriented features (line, edge fragments) are nearly collinear, then that is an *unexplained coincidence* if they are seen as unrelated to each other, but is *explained* if they are interpreted as part of the same contour. Likewise, if two localized features (dots, end points) fall nearby each other, then that is an *unexplained coincidence* if they are represented as unrelated to each other, but is *explained* if they are represented as part of the same object. In computational experiments, the predicate `coincident` has been defined using a simple distance threshold. This is almost certainly not a psychologically valid definition, in part because it is not scale-invariant, but it gives reasonable results in dot grouping (e.g., Fig. 7a). Other definitions, e.g., distance threshold scaled by local item density, may also be used. In any event, the psychological strength of the proximity cue is something to be discovered by experiment (e.g., see Kubovy, Holcombe & Wagemans 1998; Kubovy & Wagemans, 1995). The main issue in MM theory is rather the logic whereby this local decision is combined with others throughout the image to form a global percept.

Various regularities (e.g., collinearity, coplanarity, symmetry, etc.) are not all equally perceptually salient, and one might well imagine that some sort of system of weights or priorities attached to the various regularities might help to fine–tune the minimization rule and better match human intuitions. Indeed, this idea may be worth pursuing, but there are several reasons to hesitate. First, it should be noted that reasonably good results (e.g., Fig. 7) have been obtained without any weights. While adding weights would no doubt improve the fit to empirical data, the improvement would come at the cost of many added parameters in the model. Second, there is an argument (admittedly not a completely compelling one) that certain regularities ought to have equal inferential strength, e.g., those that involve accidents of exactly one degree of freedom. For example, collinearity and parallelism each involve zero difference in orientation (a coincidence in one df), while coplanarity between an end point and a plane involves zero distance (again a coincidence in one df). Conversely, coincidences of *more* than one degree of freedom (e.g., cotermination of two endpoints in the plane, 2 dfs) are "more surprising" accidents and hence should carry more inferential weight. [4] This argument suggests that the "right"

[4] The number of dimensions involved in the coincidence is the *codimension*; see previous footnote.

weighting is one in which each regularity predicate expresses a one-df accident, with no additional prioritization. This argument is not completely satisfying in part because it ignores the possibility that different parameters might have different probability distributions. For example, while collinearity and perpendicularity are both regularities along an angle parameter, the parameter might have a tighter distribution around collinear than around 90°. Similarly, the distribution for spatial coincidence (i.e., the basis for the judgment of proximity) might have a very wide distribution (again see Kubovy, Holcombe & Wagemans, 1998).

The maximum-depth rule is evidently related to other perceptual rules from the literature. Most obviously, it is a variant of the classic "minimum principle": choose the simplest interpretation possible (Hochberg & McAlister, 1953; Leeuwenberg, 1971; Hatfield & Epstein, 1985). As mentioned above the maximum-depth rule also optimally satisfies Rock's "coincidence explanation principle" (Rock, 1983). Historically, the minimum principle has often been counterposed with the "likelihood principle": choose the interpretation most likely to be true. The maximum-depth rule proposed here reconciles these two principles by explicitly showing how under a suitable model of minimality, the minimal model *is* the most likely to be true (cf. Chater, 1996). [5]

One advantage of MM theory over other extant proposals is that it is completely explicitly defined and fully computable. The most closely comparable proposal is structural information theory (also called coding theory) of Leeuwenberg and his colleagues (Boselie & Leeuwenberg, 1986; Buffart, Leeuwenberg & Restle, 1981; Leeuwenberg, 1971; Van der Helm & Leeuwenberg, 1991), which is in many ways similar in spirit (see Pomerantz & Kubovy, 1986 for summary and a critique). In structural information theory, contours of competing candidate interpretations are expressed as strings of symbols, and then the strings are compacted into a minimal form by explicitly encoding repetitions, symmetries, etc. The interpretation with the shortest minimal code is the winner. Structural information theory has been remarkably successful in predicting human percepts in a variety of perceptual domains. Progress has been made in finding efficient algorithms for code minimization (Van der Helm & Leeuwenberg, 1986). However, one difficulty in rendering a fully computable version of the theory is that the choice of candidate interpretations is not automatically determined; in practice the scientist must select them based on subjective inspection of the figure. Perhaps for this reason, structural information theory has not been taken up by researchers in computer vision. A more abstract problem with the theory concerns the motivation behind its elemental operations; as with

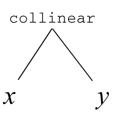
[5] That is, in many treatments pitting the likelihood and minimum principles against each other, it is assumed that the two principles point to *different* perceptual interpretations – and, indeed, depending on how the principles are cashed out, they may well do so. By contrast MM theory specifies a minimality criterion that can be shown to systematically pick out the interpretation most likely to be veridical. Hence, both principles are satisfied, and the question of which principle is primary becomes moot. Such a link between minimality and likelihood has been explicitly called for in the literature (Pomerantz & Kubovy, 1986).

Gestalt laws, it is not clear exactly why the basic minimization rules in the theory work (though see Van der Helm & Leeuwenberg, 1991 for steps in this direction).

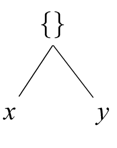
From a certain point of view, MM theory is agnostic about the choice of regularities used in the construction of interpretations. Indeed, no a priori basis exists for this choice. Rather the thrust of the theory is to suggest a *rational mapping* from a given choice of regularities to a given interpretation; that is, a systematic way of deciding which interpretation one should choose *given* a certain choice of assumptions. However, it can be shown that the choice of regularities is determined by a choice of grouping units – i.e., a choice of which types of aggregations the observers would like to parse the world into. Again here no a priori basis for choice exists. But a commitment to (say) contours and surfaces seems somehow less tendentious than one to (say) collinearity and coplanarity; and in any case, some commitment of the former type is made by *every* perceptual theory. The issue of how grouping units give rise to particular preferences is thus a crucial one, and leads eventually to proposal for a formal definition of the most abstract grouping unit, "object". Hence, the next section spells out in more detail how the observer's assumptions about grouping units lead to interpretive preferences in the theory.

5. Existential axioms

Consider the simple configuration of two collinear line segments in Fig. 8a. This configuration is described by either of two simple trees.



or



The first tree is more regular than the second tree, which omits any mention of the regularity present in the configuration. Hence, the first tree is below the second in the $\leqslant$-ordering, and the interpretation space is as shown in Fig. 8. Consequently the first tree is minimal and is preferred. Again, the logic here is that the observed collinearity between the two line segments is left unexplained – a suspicious coincidence – by the non-preferred tree, but is generic, and thus *explained*, under the preferred tree.

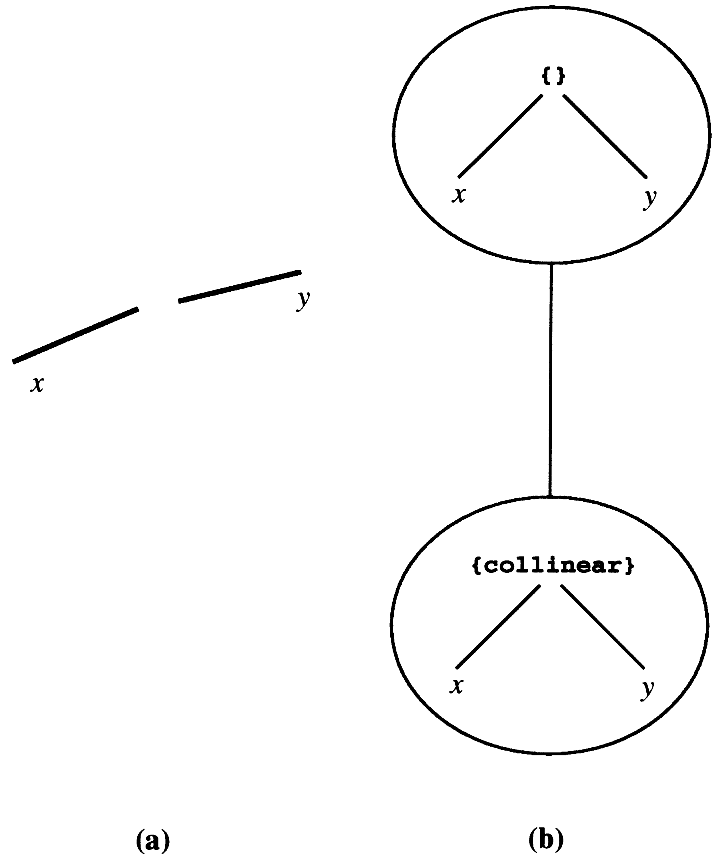


Fig. 8. (a) A configuration of two line segments and (b) the interpretation space for this figure, showing the partial order. The lower tree is minimal and is preferred.

Notice how the observer's tacit belief that *there exist contours in the environment* licenses this ordering. Under the hypothesis that the two line segments are part of a contour, the collinearity is generic – expected – and hence explained. Under the "null" hypothesis of no structure, the collinearity would remain an unexplained coincidence. If contours do not exist in the environment, then there is no model to explain the collinearity, and hence collinearity is a meaningless spatial configuration, no more special then a 37° angle. If contours do exist, then collinearity becomes something which sustains a stable explanation, and hence should be noted when present.

Fig. 9 gives examples of how certain characteristic regularities correspond to certain stereotypical structures in the world. Just as contours generically give rise to collinearities, surfaces generically give rise to certain patterns of regularities. For example, the occluding edges of a surface will tend to be consecutively coterminous (Fig. 9b) or coplanar (Fig. 9c). (In the figure, parallel line segments are interpreted

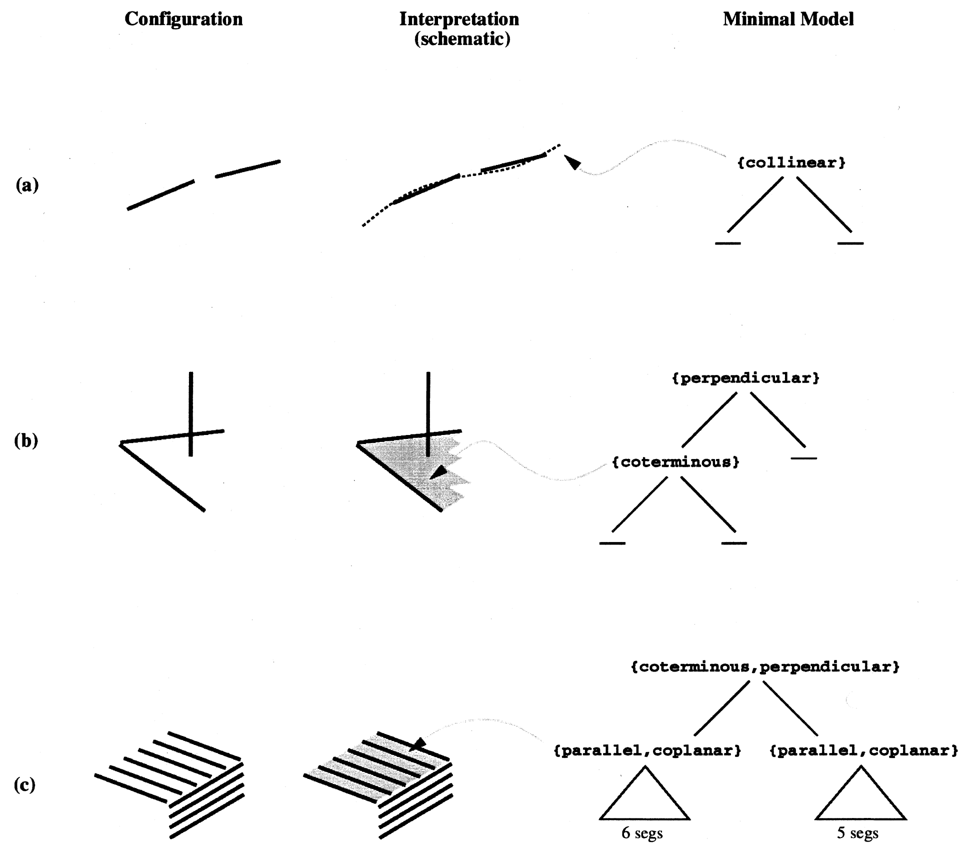


Fig. 9. Figure showing how certain types of grouping units lead to the expectation of certain patterns of regularities. In (a) a `collinear` term corresponds to a contour. In (b) and (c) various combinations of regularity terms correspond to surfaces.

non-accidentally as parallel in 3D and hence as coplanar.) Again, each of these special configurations is neatly explained by the corresponding world structure, so that interpretations containing this structure are to be preferred.

Putting this more formally, we can envision *existential hypotheses* of the form:

Existential axioms

Axiom 2a. There exist *contours.*

Axiom 2b. There exist *surfaces.*

These axioms cash out Axiom 2 (scene primitives) from Section 3. Of course, the existence of contours and surfaces is a very basic physical fact of our environment, the converse of which is hard to imagine. But this certainly does not diminish the importance of an observer's assuming it to be true; indeed this is exactly why the system *must* assume it in order for its interpretations to come out right (Feldman, 1998).

From a logical point of view, the preference order is produced by the entailment

$$\exists contours \Rightarrow [\{\texttt{collinear}\} \leqslant \{\}],$$

from which the partial order shown in Fig. 8 follows. Similarly, for surfaces,

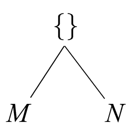
$$\exists surfaces \Rightarrow [\{\texttt{coterminous}\} \leqslant \{\}],$$
$$\exists surfaces \Rightarrow [\{\texttt{coplanar}\} \leqslant \{\}],$$

etc. The importance of these hypotheses is a logical analog of the fact that, contrary to what appears to be a widespread belief in the literature, conventional non-accidental inference is not justified from a Bayesian point of view unless the associated world structure (e.g., in the case of 2D parallelism, 3D parallelism) is assumed to have elevated prior probability in the world (Jepson & Richards, 1992).

Many details of the logical formalism have been omitted here for ease of exposition. The geometry of curves and surfaces leads to the prominence of regularities other than those discussed here, and in any case contours and surfaces are not exhaustive of sub-object-level units. In a more general sense, it is possible to prove that for any image x and class of grouping units P, the minimal model will be the *correct* parse – will correctly recover the original primitives – with probability arbitrarily close to one. (The possibility of error here derives from the small but ever-present possibility that regularities are satisfied by accident, leading to accidental interpretations.) This fact resolves the mystery of why more regular interpretations are preferred: if the partial order is induced in a suitable way, then the most regular interpretation is usually the *objectively correct* interpretation. We now attempt to extend the above argument to objects in general.

6. Objects

Treisman (1986) refers to objects as "complex wholes". The object concept proposed here formalizes both the notion of "complex" and the notion of "whole". Consider again the parse trees in Fig. 7. In (a) and (b), the configurations consist phenomenally of two objects, with only generic or "irregular" structure between them. Correspondingly, the head term of each minimal parse is the generic term { }, while each of the individual objects is a subtree hanging from the generic term. In general, one might imagine that two objects in a generic relationship – i.e., independent and with no special juxtaposition – would have a minimal parse with this form. Hence, we postulate that in a tree of the form



M and N are both "objects", provided that $depth(M) > 0$ and $depth(N) > 0$. We need this provision because otherwise M and N could simply be empty models themselves, containing no internal structure at all (perhaps even being head terms to yet more generic structure beneath them – which is how random noise such as the background in Fig. 7c is represented).

This leads immediately to a very straightforward definition for objects. First, for any node M in a parse tree, denote by $\pi(M)$ the parent of M (that is, the node from which it hangs; note that this is always unique). If M is the top node in the tree, then we say M has no parent. For any tree T, denote by $h(T)$ the head term of T. Then the following definition seems to capture what we mean by an object in the context of grouping.

Definition (*Object*). In a configuration whose minimal interpretation is T, a subtree Q appearing in T is an object if

$$\text{(a) } \pi[h(Q)] = \{\}, \text{ or } h(Q) \text{ has no parent, and}$$
$$\text{(b) } depth[h(Q)] > 0.$$

Loosely, part (a) of the definition means that Q is a "whole", and part (b) means that Q is "complex". More specifically, part (a) means that Q is not simply the subtree of a larger, more complex object; rather it goes all the way to the top of the tree describing the structure within it – i.e., it is the whole. Part (b) means that Q contains some regular structure, and is not simply a random juxtaposition of visual items. Nevertheless, in keeping with the very fluid notion of object, this definition does not say *what* structure Q must contain: any type of regularity will do.

Fig. 2c shown above gives some examples. In each of these configurations, the minimal interpretation has a positive-depth term in its head, thus satisfying the definition. More importantly, if any two of these objects are juxtaposed in one scene *at random*, the minimal interpretation would then have { } in its head and the two objects as subtrees (with probability arbitrarily close to one). On the other hand, if two objects are "glued together" in some non-accidental way, e.g., two surfaces placed perpendicularly, then the resulting scene would be cognized as a single object, as predicted (cf. Leeuwenberg & Van der Helm, 1991).

The main argument in favor of this definition is *prima facie*. Objects are entities in the scene within which there is a great deal of regular structure, and between which there is little or none. When two objects act as if their structures and locations are very coordinated, like two collinear lines or two inter-locked puzzle pieces, then they become phenomenally one object.

Evidence from the literature supports the idea that items that appear to obey a mutually constraining regularity are perceived as one object. Wagemans, Van Gool, Swinnen and Van Horebeek (1993) and Feldman (1997a) found that dots tend to cohere into a single unit the more regularity (symmetry, collinearity, etc.) is perceived among them. Similarly, Pomerantz and Pristach (1989) argue that grouping is driven by the extraction of "emergent features" from collections of items rather than by simple aggregation. Working in the context of structural information theory, Van

Lier, Van der Helm and Leeuwenberg (1994) emphasize the contrast between *internal structure* and *external structure*, analogous to the dichotomy here between a given node and its parent in the overall tree.

The proposed object concept dictates that the observer should divide up inferred minimal parse trees at depth-0 nodes. The assumption is that the resulting component subtrees – objects – will tend to exhibit the various extra-visual object properties discussed by Spelke and others: persisting over time, mediating causality, and so forth. It is extremely non-obvious that *any* consistent way of dividing up parse trees will tend to create subtrees with such remarkable properties. The hypothesis that there exists such a way, like the assumption that contours and surfaces exist, is a very basic assumption that the observer must make in order to make semantic sense of the visual world. This is the *object hypothesis*.

Axiom 2c. *There exist objects* (compare Gregory, 1970).

This hypothesis complements Axioms 2a and 2b in cashing out Axiom 2 (scene primitives). Without this axiom, the object concept proposed here would still be well-defined from a syntactic point of view, but it would be pointless semantically.

The proposed object definition describes "perfect" objects: all structure within the object, no structure between objects. In practice one might imagine that two juxtaposed objects, even if the juxtaposition was "regular", might still tend to seem like two attached objects (rather than one object) if the structure within each object was much greater than the structure joining them (see Van Lier et al., 1994). Formally, this suggests that the degree of objecthood of a subtree might be related to the *difference* between the depth of its head term and that of its head term's parent. This notion is easy to formalize.

Proposal (degree-of-objecthood)**.** For any tree Q whose head term is $h(Q)$, the degree of subjective objecthood of Q will increase monotonically with the numeric quantity

$$\lambda(Q) = depth[h(Q)] - depth[\pi[h(Q)]].$$

We choose the difference here rather than some other contrast measure (e.g., the ratio) because the two depths can be interpreted as dimensionalities of vector spaces (see Feldman (1997c)), so their difference can also be interpreted as the dimension of some space, [6] whereas the ratio is a meaningless quantity.

$\lambda(Q)$ is a measure of how much structure is interior to an object compared to how much structure binds it to other components of the scene. The higher the $\lambda(Q)$ the more the structure is inside the object, and hence the more phenomenally object-like we expect Q to be. This proposal is speculative, and empirical data should be addressed to it; but Fig. 10 suggests that it is approximately right. The proposal also explains the mixed intuition about, say, a nose on a face: a nose's high degree of

[6] See Poston and Stewart (1978) for discussion of the meaning of such dimensional differences.

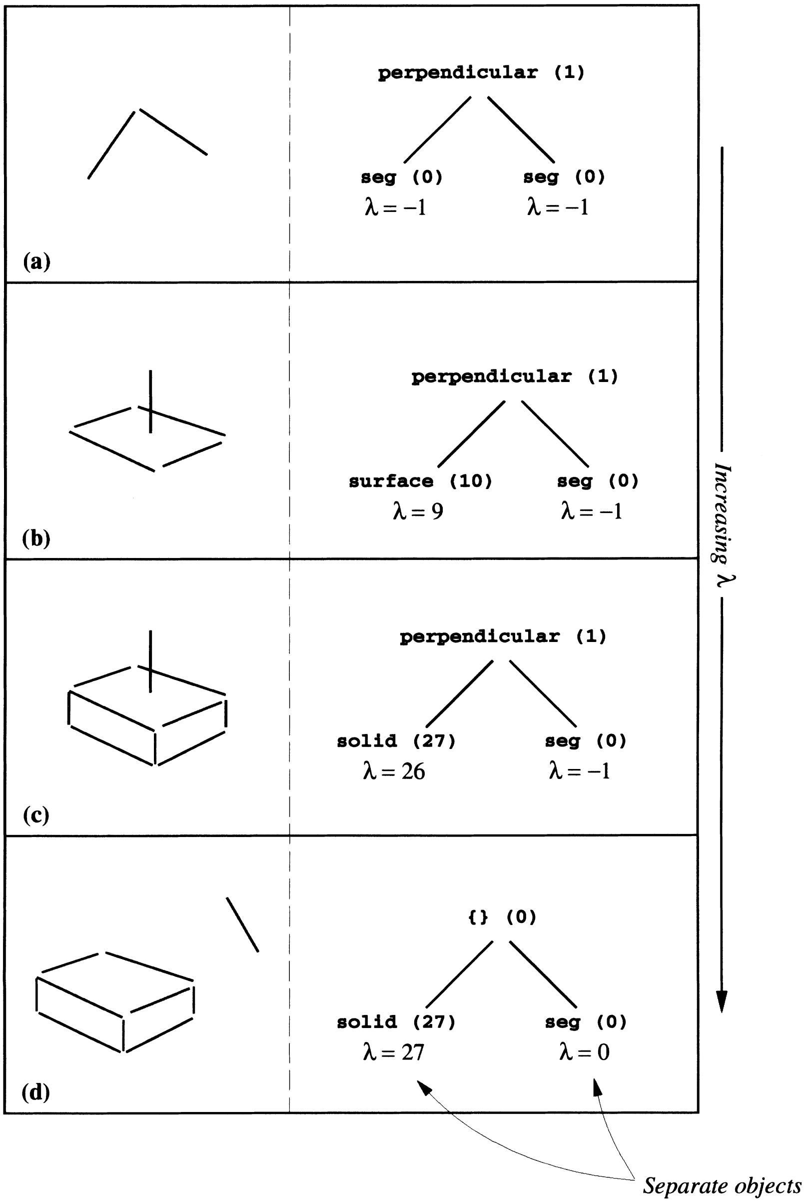


Fig. 10. showing change in $\lambda(Q)$ with changing scene structure. Models are given in abbreviated notation, with depth given in parentheses. For each model, λ is the difference between its depth and that of its parent. In (a) the two segments clearly form one object. In (b) there is more of a sense of two components, but still one object. In (c) this sense diminishes as one part becomes more internally structured. In (d) there is no structure binding the two parts at all (i.e., the depth of the head term is now zero) and the scene is perceived as two objects.

internal structure suggests that it is a separate object from the face, but its fairly regular attachment to the face (perpendicular, coplanar, and dynamically tethered) suggests that it is simply a part of a larger object (the head). A similar mixed case for which the definition gives a reasonable answer is that of houses in a row (Fig. 11). If each house is an internally unstructured configuration, like a dot (perhaps as seen from a distance), then the *row* becomes a phenomenal object. When some internal structure within each house comes into view, while the regularity between houses does not change, λ for each house increases and the houses begin to be seen as separate objects. Finally when only a small number of houses are visible, and the

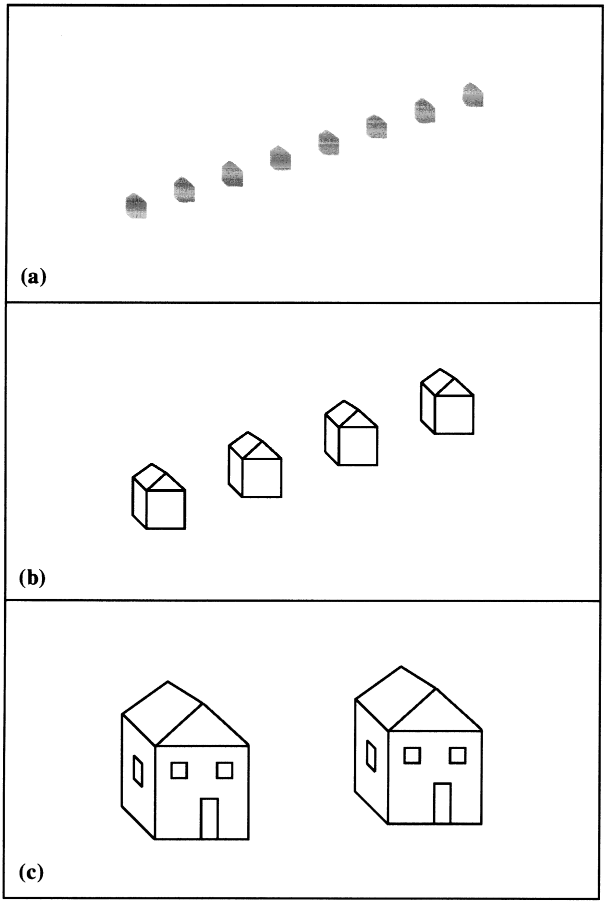


Fig. 11. Zooming in on a row of houses. (a) Seen from a distance, internal details of each house are not visible, while the collinearity in the entire row of houses is salient. Hence, the *row* is perceived as a single object, of which the individual houses are parts. (b) At a closer viewing distance, when some internal details of each house becomes visible (leading to higher λ for each house) the houses begin to seem like individual objects, though still hierarchically part of the row. (c) When only two houses are visible, each with much internal detail visible, the regularity in the configuration ceases to be represented, and each house is a single object.

collinearity between adjacent houses ceases to be perceptually represented, each house becomes a perfectly coherent object unto itself.

Note that this proposal does not mean that there are no spatial relations between objects. Rather it means that there are no *non-accidental* spatial relations between objects. Parts of the scene bound by a non-accidental relationship will be cognized as a single object. When two otherwise distinct objects move in tandem, like a horse and rider, they tend to be regarded as a single object (albeit one with two well-defined parts). Again the degree-of-objecthood proposal predicts that such an object will seem relatively weak, and its constituent parts (the horse and rider) relatively strong – which seems psychologically correct.

7. Conclusion

In Cognitive science, one learns to be suspicious of definitions. Too often they stipulate that which was supposed to be explained, sweeping the subtleties of the phenomenon under the rug. The object definition proposed here is not intended to replace human intuitions; quite the contrary, it attempts to describe them, and hence defers to them. Clearly, much remains to be worked out and specified. Yet in order to elevate objects above the level of "I know one when I see one", one needs to commit to a formalism as a starting point for a more rigorous discussion.

The main goal of the object definition – like MM theory in general – is to render human intuitions in mathematical form. The maximum-depth rule bears an appealing resemblance to the informal notion of "inference to the best explanation": the minimal interpretation is the image description that explains the most of what seems to need to be explained about the image. The "object hypothesis" amounts to the belief that the best explanation of the image will sometimes tend to be compartmentalized into distinct coherent bundles. The proposed object definition simply attaches a concrete meaning to this idea. One measure of the power of a theory is its ability to solve pertinent problems that it was not specifically designed to solve. If the object definition deriving from minimal model theory turns out to be useful, it tends to lend credibility to the theory.

References

Ashby, F. G., Prinzmetal, W., Ivry, R., & Maddox, W. T. (1996). A formal theory of feature binding in object perception. *Psychological Review*, *103*, 165–192.

Baillargeon, R. (1994). How do infants learn about the physical world?. *Current Directions in Psychological Science*, *3*(5), 133–140.

Barchilon Ben-Av, M., Sagi, D., & Braun, J. (1992). Visual attention and perceptual grouping. *Perception and Psychophysics*, *52*(3), 277–294.

Barchilon Ben-Av, M., & Sagi, D. (1995). Perceptual grouping by similarity and proximity: experimental results can be predicted by intensity autocorrelations. *Vision Research*, *35*(6), 853–866.

Barlow, H. (1994). What is the computational goal of the neocortex? In C. Koch & J. L. Davis (Eds.), *Large-scale neuronal theories of the brain* (pp. 1–22). Cambridge: MIT Press.

Barrow, H. G., & Tenenbaum, J. M. (1981). Interpreting line drawings as three-dimensional surfaces. *Artificial Intelligence*, *17*, 75–116.

Baylis, G., & Driver, J. (1993). Visual attention and objects: Evidence for hierarchical coding of location. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(3), 451–470.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115–147.

Binford, T. (1981). Inferring surfaces from images. *Artificial Intelligence*, *17*, 205–244.

Boselie, F., & Leeuwenberg, E. (1986). A test of the minimum principle requires a perceptual coding system. *Perception*, *15*, 331–354.

Buffart, H., Leuwenberg, E., & Restle, F. (1981). Coding theory of visual pattern completion. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(2), 241–274.

Caelli, T. M., & Umansky, J. (1976). Interpolation in the visual system. *Vision Research*, *16*(10), 1055–1060.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review*, *103*(3), 566–581.

Clowes, M. B. (1971). On seeing things. *Artificial Intelligence*, *2*, 79–116.

Davey, B., & Priestley, H. (1990). *Introduction to lattices and order*. Cambridge: Cambridge University Press.

Feldman, J. (1996). Regularity vs genericity in the perception of collinearity, *Perception 25* 335–342.

Feldman, J. (1997a). Curvilinearity, covariance, and regularity in perceptual groups. *Vision Research*, *37*(20), 2835–2848.

Feldman, J. (1997b) Efficient regularity-based grouping. In *Proceedings of the I.E.E.E. conference on computer vision and pattern recognition* (pp. 288–294). Los Alamitos: Computer Society Press.

Feldman, J. (1997c). Regularity-based perceptual grouping. *Computational Intelligence*, *13*(4), 582–623.

Feldman, J. (1997d). The structure of perceptual categories. *Journal of Mathematical Psychology*, *41*, 145–170.

Feldman, J. (1998). Does vision work? Towards a semantics of perception. In E. Lepore & Z. Pylyshyn (Eds.), *What is cognitive science?* New York: Basil Blackwell (forthcoming).

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gilchrist, A. L. (1977). Perceived lightness depends on perceived spatial arrangement. *Science*, *195*, 185–187.

Glass, L. (1969). Moiré effects from random dots. *Nature*, *223*, 578–580.

Gregory, R. L. (1970). *The intelligent eye*. New York: McGraw-Hill.

Guy, G., & Medioni, G. (1996). Inferring global perceptual contours from local features. *International Journal of Computer Vision*, *20*(1/2), 113–133.

Hatfield, G., & Epstein, W. (1985). The status of the minimum principle in the theoretical analysis of visual perception. *Psychological Bulletin*, *97*(2), 155–186.

He, Z. J., & Nakayama, K. (1992). Surfaces versus features in visual search. *Nature*, *359*, 231–233.

Hochberg, J., & McAlister, E. (1953). A quantitative approach to figural "goodness". *Journal of Experimental Psychology*, *46*, 361–364.

Hock, H. S., Tromley, C., & Polmann, L. (1988). Perceptual units in the acquisition of visual categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 75–84.

Jacobs, D. (1996). Robust and efficient detection of salient convex groups. *I.E.E.E. Transactions of Pattern Analysis and Machine Intelligence*, *18*, 23–37.

Jepson, A. & Richards, W. A. (1992). What makes a good feature? In L. Harris & M. Jenkin (Eds.), *Spatial vision in humans and robots* (pp. 89–125). Cambridge: Cambridge University Press.

Kahneman, D. & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization* (pp. 181–211). Hillsdale, NJ: Lawrence Elrbaum.

Kanizsa, G. (1979). *Organization in vision: essays on Gestalt perception*. New York: Praeger.

Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, *23*, 141–221.

Kramer, A., & Jacobson, A. (1991). Perceptual organization and focused attention: The role of objects and proximity in visual processing. *Perception and Psychophyics*, *50*(3), 267–284.

Kubovy, M., Holcombe, A. O., & Wagemans, J. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology*, *35*, 71–98.

Kubovy, M., & Wagemans, J. (1995). Grouping by proximity and multistability in dot lattices: a quantitative gestalt theory. *Psychological Science*, *6*(4), 225–234.

Lawson, R. (1999). Achieving visual object constancy over plane rotation and depth rotation. *Acta Psychologica*, 221–245.

Leeuwenberg, E. (1971). A perceptual coding language for visual and auditory patterns. *American Journal of Psychology*, *84*(3), 307–349.

Leeuwenberg, E., & Van der Helm, P. (1991). Unity and variety in visual form. *Perception*, *20*, 595–622.

Link, N. K., & Zucker, S. W. (1987). Sensitivity to corners of flow patterns. *Spatial Vision*, *2*(3), 233–244.

Lowe, D. G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, *31*, 355–395.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London B*, *200*, 269–294.

Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science*, *257*, 1357–1363.

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, *9*, 441–474.

Palmer, S. E. (1978). Structural aspects of visual similarity. *Memory and Cognition*, *6*(2), 91–97.

Pizlo, Z., Salach-Golyska, M., & Rosenfeld, A. (1997). Curve detection in a noisy image. *Vision research*, *37*(9), 1217–1241.

Pomerantz, J. R., & Kubovy, M. (1986). Theoretical approaches to perceptual organization. In: K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance*, *vol. 2*, *Cognitive processes and performance* (pp. 36-1–36-46). New York: Wiley.

Pomerantz, J. R., & Pristach, E. A. (1989). Emergent features, attention, and perceptual glue in visual form perception. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(4), 635–649.

Poston, T., & Stewart, I. N. (1978). *Catastrophe theory and its application*. London: Pitman.

Prazdny, K. (1984). On the perception of Glass patterns. *Perception*, *13*, 469–478.

Prinzmetal, W., & Banks, W. (1977). Good continuation affects visual detection. *Perception and Psychophysics*, *21*(5), 389–395.

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, *32*, 65–97.

Quine, W. (1965). On simple theories of a complex world. In M. H. Foster & M. L. Martin (Eds.), *Probability, confirmation, and simplicity: Readings in the philosophy of inductive logic* (pp. 250–252). New York: Odyssey Press.

Reiter, R., & Mackworth, A. K. (1989). A logical framework for depiction and image interpretation. *Artificial Intelligence*, *41*, 125–155.

Richards, W. A. (1988) *Natural computation*. Cambridge, MA: MIT Press.

Richards, W. A., Jepson, A., & Feldman, J. (1996). Priors, preferences, and categorical percepts. In D. Knill & W.A. Richards (Eds.), *Perception as Bayesian inference* (pp. 93–122). Cambridge: Cambridge University Press.

Rock, I. (1983). *The logic of perception*. Cambridge: MIT Press.

Rock, I., Linnett, C. M., Grant, P., & Mack, A. (1992). Perception without attention: Results of a new method. *Cognitive Psychology*, *24*(4), 502–534.

Sekuler, A. B., Palmer, S. E., & Flynn, C. (1994). Local and global processes in visual completion. *Psychological Science*, *5*(5), 260–267.

Smits, J. T., & Vos, P. G. (1987). The perception of continuous curves in dot stimuli. *Perception*, *16*(1), 121–131.

Sober, E. (1975). *Simplicity*. London: Oxford University Press.

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, *14*, 29–56.

Stevens, K. A. (1978). Computation of locally parallel structure. *Biological Cybernetics*, *29*, 19–28.

Takeichi, H., Nakazawa, H., Murakami, I., & Shimojo, S. (1995). The theory of the curvature-constraint line for a modal completion. *Perception*, *24*(3), 373–389.

Tarr, M., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, *21*, 233–282.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 194–214.

Treisman, A. (1986). Properties, parts, and objects. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance*, *vol. 2*, *Cognitive processes and performance* (pp. 35-1–35-70). New York: Wiley.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*, 107–141.

Tse, P. (1999). Complete merge ability and amodal completion. *Acta Psychologica*, 165–201.

Ullman, S. & Shashua, A. (1988). Structural saliency: The detection of globally salient structures using a locally connected network. In *Proceedings of the second international conference on computer vision* (pp. 321–327). Tampa, FL.

Van der Helm, P. A., & Leeuwenberg, E. L. J. (1986). Avoiding explosive search in automatic selection of simplest pattern codes. *Pattern Recognition*, *19*(2), 181–191.

Van der Helm, P. A., & Leeuwenberg, E. L. J. (1991). Accessibility: A criterion for regularity and hierarchy in visual pattern codes. *Journal of Mathematical Psychology*, *35*(2), 151–213.

Van der Helm, P. A., & Leeuwenberg, E. L. J. (1996). Goodness of visual regularities: A nontransformational approach. *Psychological Review*, *103*(3), 429–456.

Van Lier, R., Van der Helm, P. A., & Leeuwenberg, E. L. J. (1994). Integrating global and local aspects of visual occlusion. *Perception*, *23*(8), 833–903.

Van Lier, R., Leeuwenberg, E. L. J., & Van der Helm, P. A. (1995). Multiple completions primed by occlusion patterns. *Perception*, *24*, 727–740.

Van Lier, R. (1999). Investigating global effects in visual occlusion: from a partly occluded square to the back of a tree-trunk. *Acta Psychologica*, 203–220.

Wagemans, J. (1992). Perceptual use of non-accidental properties. *Canadian Journal of Psychology*, *46*(2), 236–279.

Wagemans, J., Van Gool, L., Swinnen, V., & Van Horebeek, J. (1993). Higher-order structure in regularity detection. *Vision Research*, *33*(8), 1067–1088.

Xu, F. (1999). Object individuation in infancy: The role of spatiotemporal information, featural information, and language. *Acta Psychologica*, 113–136.

Zucker, S. W. (1985). Early orientation selection: tangent fields and the dimensionality of their support. *Computer Vision Graphics, and Image, Processing*, *32*, 74–103.

Zucker, S. W., Stevens, K. A., & Sander, P. (1983). The relation between proximity and brightness simillarity in dot patterns. *Perception and Psychophysics*, *34*(6), 513–522.