Appeared as: *Computational Intelligence*, **13(4)**, 582--623

Regularity-based Perceptual Grouping

Jacob Feldman
Dept. of Psychology, Center for Cognitive Science
Rutgers University

This paper investigates perceptual grouping from a *logical* point of view, defining a grouping interpretation as a particular kind of logical expression, and then developing an explicit inference theory in terms of such expressions. First, a regularity-based interpretation language is presented, in which an observed configuration is characterized in terms of the *regularities* (special configurational classes, e.g. non-accidental properties) it obeys. The most preferred interpretation in such a system is shown to be the most-regular (maximum "codimension") model the observed configuration obeys, which is also the unique model in which it is *generic* (typical). Inference then reduces to a straightforward exercise in Logic Programming. Because generic model assignment involves negation, this reduction requires that a version of the Closed World Assumption (CWA) be adopted.
Next, this entire regularity-based machinery is generalized to the grouping problem: here an interpretation is a hierarchical (recursive) version of a model called a *parse tree.* For a given number of dots and a fixed choice of regularity set, it is possible to explicitly enumerate the complete set of possible grouping interpretations, partially ordered by their degree of regularity (codimension). The most preferred interpretation is the one with maximum codimension (i.e., the most regular interpretation), which we call the *qualitative parse*. An efficient procedure (worst case $O(n^2)$) for finding the qualitative parse is presented. The qualitative parse has a unique epistemic status: given a choice of regularity set, it is the only grouping interpretation that both (a) is maximally regular, and (b) satisfies the CWA. This unique status, it is argued, accounts for the perceptually compelling quality of the qualitative parse.

Introduction: the logic of grouping

Consider the four dots in Fig. 1. It is natural for human observers to attribute some "structure" to such a configuration; one such interpretation is notated in the figure. The indicated interpretation includes some division into groups (a) and, among some of the dots, some degree of collinearity (b). The inference of this sort of structure is compelling to the eye, of course, but does not (in a modern view) appear to derive from any kind of absolute logical necessity. The underlying inference does, on the other hand, have some logical structure, which this paper investigates.

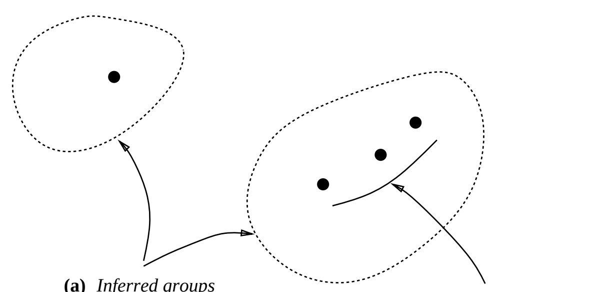


Figure 1. A field of four dots, in which a human observer might find **(a)** some grouping structure and **(b)** some (near-)collinearity structure.

The author wishes to express his gratitude to Whitman Richards, Allan Jepson, Alan Mackworth, and Ray Reiter for many helpful comments and discussions, and to three anonymous reviewers whose comments greatly improved the manuscript. This work was made possible by the Rutgers Center for Cognitive Science (RuCCS), Rutgers University, New Brunswick, New Jersey.

This interpretation can be thought of as a *model* of the dot configuration. Intriguingly, though, in the usual terminology of logicians, it is the configuration itself, rather than the interpretation, that is the "model" that instantiates a certain "theory" or logical construction. This reversal of terminology, compared to that in use among perceptual theorists, is telling. This paper works out some implications of considering scene interpretation, in particular dot grouping, from a "logical" point of view. In this approach, influenced by the framework proposed by Reiter and Mackworth (1989); see Mackworth, 1988) interpreting an observed configuration will reduce to identifying a sentence of a certain logical language that the observed configuration satisfies in a certain way. The result will be a model of the dot configuration (in this case a grouping interpretation) that uniquely satisfies a certain set of desirable logical constraints.

The process of finding groups among visual items has been regarded in a number of ways: as the result of low-level neural mechanisms (Glass, 1969; Caelli & Julesz, 1978; Prazdny, 1984); as the result of reflexive reasoning principles (e.g. Gestalt laws); or as the result of computational procedures emulating such principles (see Stevens, 1978; Jacobs, 1989; Brookes & Stevens, 1991; Zucker, 1985; Guy & Medioni, 1992; and Cox, Rehg, & Hingorani, 1993 for recent examples). All these approaches have merits, and are not necessarily to be regarded (at least, a priori) as inconsistent with one another, as they attack the problem at different levels. The logical approach presented here presents an alternative. We cast grouping interpretation into a formal language with its own rules of construction, and then consider logical constraints on these interpretations such that a single interpretation uniquely satisfies these constraints. Accordingly, computing this grouping interpretation will be treated in the context of Logic Programming, which considers how logical structures can be computed. This setting allows for such previously inaccessible issues as the "semantics" of grouping to be approached in a concrete fashion.

It turns out that there is a natural way of expressing grouping interpretations as logical structures, which will be laid out in Sections and . We begin by recapitulating some scene interpretation machinery (presented in part in Feldman (1991, 1992a, 1992b, 1997b): first informally for motivation, using an example from shape classification (Sec.), then more formally using notation that will be incorporated into the grouping theory (Sec.). The interpretation theory was originally conceived for characterizing shape classes, but grouping interpretations are nothing more than a recursive generalization of the same idea, in which each group is characterized as a certain configuration type, and then the overall interpretation is a characterization of the relationship among the various groups. Suitably formalized, this hierarchical grouping interpretation reduces to a sentence in a certain logical language. This language has a rich internal structure that is explored in detail in Section . Finding the "correct" grouping of a set of items then reduces to a straightforward Logic Programming problem, the solution of which (finding the "model" of the observed scene) corresponds to proving that the scene satisfies a certain logical expression.

Background

This section gives a capsule summary of the shape-classification work on which the current paper builds, giving an intuitive introduction to the major ideas. However, all the needed machinery will be derived from the ground up in a more formal manner starting in Section .

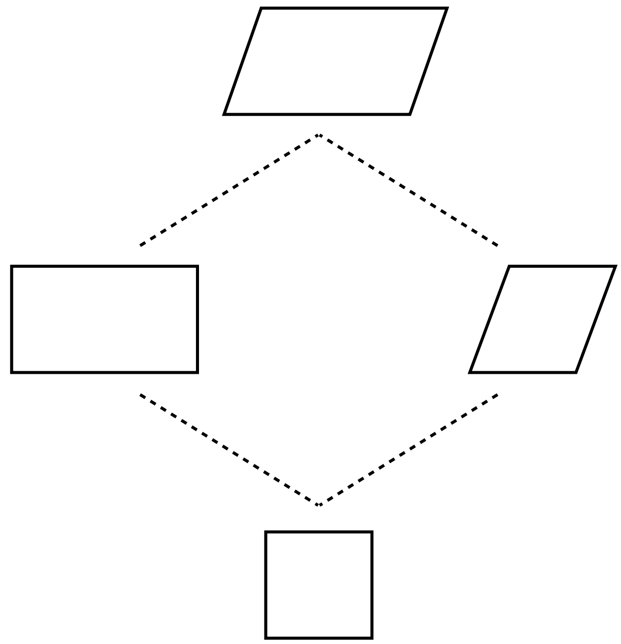


Figure 2. A regularity lattice for parallelograms and their subclasses.

Regularities and the regularity lattice. In previous work (Feldman, 1991, 1992a, 1992b, 1997b) it has been proposed that intuitively compelling shape classes can be enumerated as the nodes in a structured hierarchy termed the *regularity lattice* (Fig. 2). The idea is to model some object space (e.g., parallelograms) by a set of generative operations (e.g., stretch and skew) applied to some simple object (e.g., a square) (an idea proposed by Leyton, 1984, 1988; see also Leyton, 1992). Taken all together, the ensemble of operations parameterize the entire parallelograms space. But taken in subsets, each set of operations defines a subclass that is *qualitatively* distinct in that each picks out a class of objects that, under the underlying parameterization, share a distinct generative history. These special subclasses are enumerated and partially ordered by the regularity lattice. This enumeration is been investigated from the point of view of inductive categorization (Feldman, 1992b, 1997b) recovery of stable scene interpretations (Richards, Jepson, & Feldman, 1996), and the distribution of probability mass in a shape space (Feldman, 1996).

Notice that the special subclasses in the lattice in Fig. 2—the nodes below the top node which denotes the completely "generic" class corresponding to the entire shape space—can be regarded as *regularities*: that is, constraints that disqualify objects that obey them from being regarded as typical. For example, the leftmost node has right angles, a highly special configuration. The rightmost node has equal-length sides. The entire lattice is constructed from these two regularity concepts, with shapes assigned to nodes (classes) by evaluating what set of regularities they obey. The representation is "qualitative" in that shapes are *only* characterized in terms of their regularities, collapsing over finer distinctions among shapes that obey the same set of regularities. Hence the characterization divides the shape space into the equivalence classes, isomorphic to the nodes on the regularity lattice, and each node on the lattice can be thought of as a distinct *qualitative prototype* for the shapes in its equivalence class. The lattice thus completely enumerates the possible forms in the space that are qualitatively distinct modulo the chosen set of regularities.

The Genericity Constraint. The critical formal constraint that makes inferences possible in the above framework is the idea of *genericity:* that objects should only be associated with a prototype in which they are formally typical. This constraint rules out models (using the term in the perception-theoretic sense) in which the observed object would have to be regarded as a coincidence. For example, any particular square satisfies the requirements of being a rectangle (it falls in the rectangle class) but it does so non-generically, because it is atypical for rectangles to have equal-length sides. On the other hand, the square falls in the "square" class generically, because it is perfectly typical for *squares* to have equal-length sides. Under a particular formal definition of genericity, the generic model choice turns out to be unique. This paper will introduce some new technical results about model assignment under the Genericity Constraint in Section . These results will be necessary in building towards the main goal of the current paper, the application of the regularity-based machinery to the problem of grouping. This will only be taken up directly in Section .

Preview of the paper. The paper will be structured as follows. In Section , we formalize the idea of the recovery of the maximally regular interpretation, constructing a logical language in which each model is a valid expression. We prove some novel formal results about the structure of the space of these expressions; the main result is that it is a distributive lattice, called the *lattice of models*. In order actually to compute the appropriate model for an observed configuration, the observer must decide which model can be "proved" from it. Hence it is natural to place this machinery into the context of Logic Programming, which is concerned with how logical expressions can be evaluated computationally. Because the expression for the most regular interpretation involve *negation*, a problematic issue in Logic Programming, we must delve a bit into semantics; it turns out to be necessary to adopt an analog of the well-known "Closed World Assumption." to justify the use of "negation as failure." Section applies the regularity-based interpretation machinery to the grouping problem, introducing a hierarchical generalization of the most regular interpretation called the *qualitative parse*. The qualitative parse amounts to an explicit logical solution to the grouping problem, with dot configurations assigned only to parses in which they are generic. The treatment of grouping will present formal results corresponding to (and building on) each of the pieces of the ordinary interpretation machinery: a logical language in which interpretations are expressed, a rule for choosing the best interpretation, and finally a complete characterization of the structure of the space of qualitative parses (i.e. the grouping analog of the lattice of models). Remarkably, this turns out to be a set of disjoint set of distributive lattices. For a given configuration the qualitative parse is locally unique, and globally so except in unusual circumstances: that is, for each configuration of dots, there is one and only one grouping interpretation that simultaneously (a) is maximally regular and (b) satisfies the Closed World Assumption.

Interpretation as regularity-finding

Regularities and models

Regularities. Say we have some input configuration x chosen from a data space X. Certain types of configurations are of special interest to the observer, in that they contain some kind of suspiciously organized structure: two line segments that share an endpoint, three dots that are collinear, and so forth. We call such properties "regularities."

More formally, we regard a "regularity" R as a logical predicate defined on X, which holds on some subset $X_R \subset X$. The essential condition for R to be regarded as a "regularity" is the following *primitive preference principle*:

(1) For a configuration obeying R, prefer a model including R to one not including R (all else being equal).

That is, a regularity is a class of configurations that an observer tends to utilize or recognize when it occurs. The justification for this tendency has been ascribed by perceptual theorists to a a number of sources for particular regularities that exhibit it. For a certain class of features, Binford (1981) and Lowe (1987) attribute the preference tendency to "non-accidentalness," i.e. probabilistic reliability due to viewing geometry. Such features are special in that they are unlikely to occur in 2-D unless they occur in 3-D, leading to a high likelihood ratio in favor of inferring the feature. This probabilistic relationship has been generalized by Bennett, Hoffman, and Prakash (1989), who introduced a suitable measure-theoretic characterization of the unlikelihood of a false inference. In several articles, Jepson, Richards and Feldman (Jepson & Richards, 1992b, 1991, 1992a; Feldman, 1991, 1992a, 1992b, 1997b; Richards et al., 1996) have advocated a more "structural" view, in which more regular configurations are constructed from more generic ones by removing degrees of freedom one by one. In this case, more preferred configurations have lower dimension (higher codimension[1]), and less preferred configurations have higher dimension (lower codimension), thus satisfying Bennett *et al*'s measure condition. Each of these approaches attempts to elucidate why certain configurations obey the above preference

[1] Codimension is the difference in dimension between a geometric object and the overall space in which it is embedded (see Poston and Stewart (1978)), which corresponds here to the completely generic configuration.

condition; in the current paper we treat this as simply a primitive, calling any configuration which obeys the principle a "regularity," and then investigate the logic of the resulting inference structure.

Defn. 1 (Regularity) *A regularity R is a logical predicate defined on X and obeying (1). When R holds on $x \in X$ we say $R(x)$, otherwise $\neg R(x)$. The region $\{x \in X | R(x)\}$ is denoted X_R.*

Naturally, other conditions are required in order for a given regularity to be useful in practice as a building-block for scene interpretations; see Jepson and Richards (1992a) and Richards et al. (1996).

The regularity set. Now, assume that the observer has at hand some distinguished *set* of regularity types

(2) $$\mathcal{R} = \{R_1, R_2, \ldots, R_k\},$$

called the *regularity set*. A given configuration may satisfy some subset of these; the *larger* this subset, the *smaller* the corresponding region of X satisfying it. It may be that the various R_i's are all independent, meaning that configurations exist that satisfy any subset of $\mathcal{R}$. To be more general, we allow that there may be pairwise constraints ω of the form

(3) $$R_i \stackrel{\omega}{\rightarrow} R_j,$$

meaning that R_i can only hold if R_j does as well, i.e.

(4) $$R_i \stackrel{\omega}{\rightarrow} R_j \text{ iff } X_{R_i} \subset X_{R_j}.$$

The pairwise constraints ω can be regarded as imposing a nested structure on the various regularity regions. For brevity a regularity set coupled with an implication set will be referred to as a *context* $C = \langle \mathcal{R}, \omega \rangle$.

Models. For any set $M \subset \mathcal{R}$, the *closure* of M under ω is the set augmented by any additional regularities implied by transitive closure of $\stackrel{\omega}{\rightarrow}$. A set of regularities is called *closed* if it is the closure of some set of regularities.

Any set of regularities $\{R_1, R_2, \ldots, R_c\} \subset \mathcal{R}$ can be identified with a logical $\wedge$-expression

(5) $$R_1 \wedge R_2 \wedge \cdots \wedge R_c,$$

which is satisfied if and only if each of the regularities in the set is satisfied. We ignore the order of terms in such an expression, so that sets of regularities can be spoken of interchangeably with the corresponding $\wedge$-expressions. Each closed subset of the regularity set is called a *model*:[2]

Defn. 2 (Model) *Given a context C, any closed set of regularities $\mathcal{M} \subset \mathcal{R}$, or the corresponding $\wedge$-expression, is called a model.*

It is convenient to use the term "model" to refer to both the closed set of regularities and the corresponding $\wedge$-expression, so that we may speak of a configuration "satisfying a model" (thinking of the model as a logical expression) and also speak of "a subset of a model" (thinking of the set). Some sample contexts are given in Table 1.

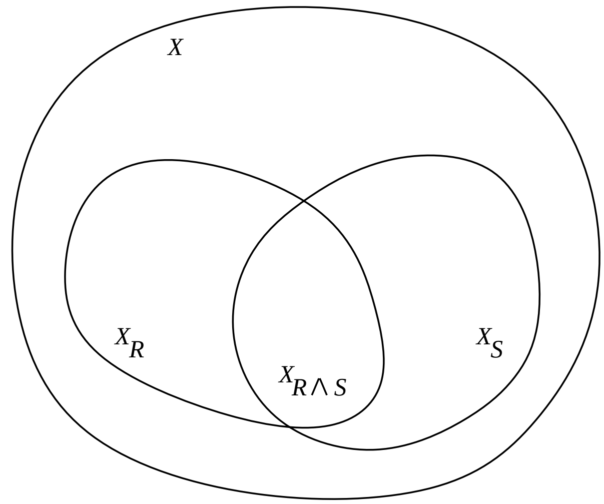


Figure 3. A Venn diagram of a configuration space X with two regularities, R and S. While some points in the space satisfy more than one model (i.e. are contained in more than one model region), each point satisfies exactly one model generically.

Notice that $\emptyset$ is a model in any context, because it is the closure of itself regardless of ω. While for a given context there are 2^k sets of regularities ($k = |\mathcal{R}|$), for non-empty ω there are fewer than 2^k distinct models. Notice that not all of the subsets of $\mathcal{R}$ correspond to distinct regions of X, but each *model* M does correspond to a distinct region $X_{\mathcal{M}}$, called the "model region."

In summary, the context $C = \langle \mathcal{R}, \omega \rangle$ entails a fixed, finite set of models $\mathcal{M}$, each of which corresponds to a distinct region $X_M \subset X$. In each context, some of the models are less constrained (contain fewer regularities), and hence correspond to large regions of X; some are more constrained and correspond to smaller regions. The least constrained (most "generic") is always $\emptyset$, because $X_\emptyset = X$. The models represent each and every distinct "way of being regular" modulo the context. That is, each model describes a configuration class that is *qualitatively distinct* from each other class modulo the context, in the literal sense that it contains different regularities.

Genericity and negation

In general, if a model $\mathcal{M}$ holds on a certain configuration x, then so do all submodels $\mathcal{M}_i \subseteq \mathcal{M}$, including $\emptyset$, which holds on all $x \in X$. The situation is depicted in Fig. 3. The "map" of the configuration space that is entailed by a given context is characterized by this nest of overlapping and sometimes nested regions. Notice that while these regions are closed under intersection, they are *not* generally closed under union. In Fig. 3, for example, $X_R \cap X_S = X_{R \wedge S}$, but $X_R \cup X_S$ is not the region for any model.

[2] Notice again that we are using the term "model" in the perception-theorist's rather than the logician's sense.

Regularities R	Implications ω	Models
$\{R\}$	-	$\emptyset, \{R\}$
$\{R,S\}$	-	$\emptyset, \{R\}, \{S\}, \{R,S\}$
$\{R,S,T\}$	$T \overset{\omega}{\to} S$	$\emptyset, \{R\}, \{S\}, \{R,S\}, \{S,T\}, \{R,S,T\}$
$\{R,S,T\}$	$T \overset{\omega}{\to} S, S \overset{\omega}{\to} R$	$\emptyset, \{R\}, \{R,S\}, \{R,S,T\}$

Table 1
Some sample contexts and their models. Models are given in set notation.

In the figure, any configuration x that satisfies a model $\{R,S\}$ (i.e. satisfies $R \wedge S$) also satisfies the models $\{R\}$, $\{S\}$, and $\emptyset$. However, as suggested above, to identify x with any of these other models would be undesirable, in that it satisfies them non-generically—specifically in that in each case x also satisfies a more regular model, namely $\{R,S\}$. This suggests the following definition of "generic."

Defn. 3 (Genericity) *For an object x and a model $\mathcal{M}$, "x satisfies $\mathcal{M}$ generically" or "x is generic in $\mathcal{M}$" iff there exists no larger model $\mathcal{M}'$, $\mathcal{M} \subsetneq \mathcal{M}'$, s.t. that x also satisfies $\mathcal{M}'$.*

Each model that a given object x obeys non-generically is missing at least one regularity that x obeys, and which would be contained in a generic model. Hence by virtue of the "primitive preference principle" (Eq. 1), it is immediate that the observer should *prefer* a generic model to any non-generic one. A non-generic model amounts to an interpretation in which the observed configuration would have to be regarded as a peculiarly atypical case. This is the so-called Genericity Constraint:

> **(Genericity Constraint):** For a configuration x, prefer a model $\mathcal{M}$ in which x is generic.

For each configuration, there is always one and only one such model.

Theorem 1 *Given a configuration x in a context C, the model $\mathcal{M}(x)$ in which x is generic exists and is unique.*

Proof. x is generic in the model $\{R|R(x)\}$, because there can be no larger models that x satisfies. Note that if x obeys no regularities in C then x is generic in the model $\emptyset$. □

The generic model is ipso facto the unique most preferred interpretation for a given configuration. The remainder of this section is devoted to characterizing this model more completely.

An immediate consequence of the uniqueness of the generic model is that the context C entails a *partition* of X, in which each cell comprises a collection of configurations all of which are generic in the same model. This partition may be viewed two different ways. In one approach, each cell in the partition is isomorphic to a model, but not identical to it, because the model region contains some points that satisfy it generically and others that satisfy it non-generically. For example, in Fig. 3, the region X_R contains both points that are generic in $\{R\}$ as well as points that are not (i.e. those that are also in X_S).

A second approach, which we pursue here, and which is essential if we are to render model assignment computable, is to attempt to give each cell in the partition an explicit logical form, so that each model spells out the points that are generic it in a completely literal way. Evidently, the key is to recognize the implicit *negative* part of each model—the part that rules out non-generic points—and make it explicit, thus completing the necessary specification of points that properly (generically) belong to the model. We call this the *completion under genericity* of the model.

First, we fix some some notation. For two sets $\mathcal{A}$ and $\mathcal{B}$, let $\mathcal{A} - \mathcal{B}$ denote the asymmetric difference $\{A \in \mathcal{A} : A \notin \mathcal{B}\}$. For any set $\{A, B, \ldots,\}$, $\overbrace{\{A,B,\ldots\}}^{+}$ means $A \wedge B \wedge \ldots$, and $\overbrace{\{A,B,\ldots\}}^{\neg}$ means $\neg A \wedge \neg B \wedge \ldots$.

Defn. 4 (Completion of a model under genericity) *For a model $\mathcal{M} \subset \mathcal{R}$, the completion under genericity is*

$$\mathcal{M}^* = \overbrace{\mathcal{M}}^{+} \wedge \overbrace{\mathcal{R} - \mathcal{M}}^{\neg}. \quad (6)$$

Because each completed model contains some number $c \leq k$ of regularities in its "positive part" and $k - c$ in its "negative part," it is notationally convenient to renumber them and rewrite the model as

$$\mathcal{M}^* = \overbrace{\{R_1, \ldots, R_c\}}^{+} \wedge \overbrace{\{R_{c+1}, \ldots, R_k\}}^{\neg}. \quad (7)$$

Model completion is always relative to a given regularity set, but with the context fixed, models and completed models are in exact one-to-one correspondence. The number c, the cardinality of the uncompleted model, or (equivalently) the cardinality of the positive part of the corresponding completed model, is called the *codimension.* It measures numerically the degree of structure exhibited by objects satisfying the given model. The role of "logical" codimension, and its relationship to the conventional geometric definition mentioned above, will be discussed below.

A configuration x satisfies a model $\mathcal{M}$ generically if an only if it satisfies the corresponding completed model $\mathcal{M}^*$, in which case we write $\mathcal{M}^*(x)$. Hence the completed model can be thought of as the "proper" (i.e., generic) name for the

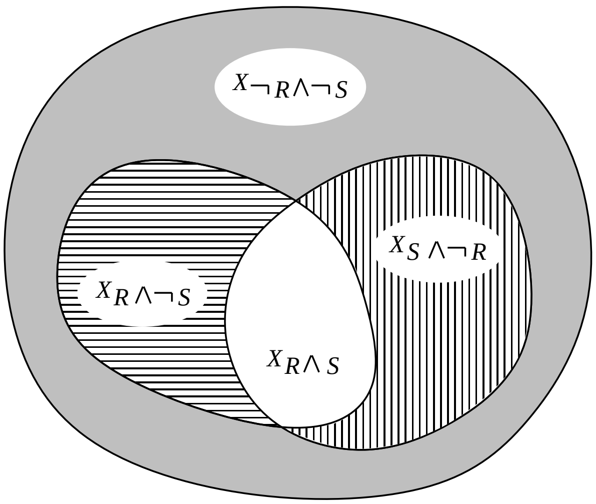


Figure 4. The same configuration space as in Fig. 3, except with each region labeled explicitly by its completed model. (Note that this is no longer a Venn diagram; the four labeled regions are disjoint.)

cell of the partition corresponding to the given model. This correspondence is pictured in Fig. 4, in which the partition is depicted and labeled explicitly.

Notice that the set of regions corresponding to completed models, which are disjoint, is not generally closed under either union or intersection.

As in the figure, note that the empty model $\emptyset$ when completed becomes $\neg R \wedge \neg S \cdots$, i.e. $\overbrace{R}^{\neg}$. To satisfy this expression, a configuration must systematically avoid satisfying any particular regularity type—a highly "special" situation. This is a natural reflection of the idea that a completely "typical" point in the configuration space is one that exhibits *no* special structure whatever.

The key point is that the completed model is a fully explicit logical form for models. A configuration x can now be assigned to a model generically simply by finding the unique completed model that succeeds on x, e.g. using Logic Programming. From this point of view it is significant that the completed model requires *negation,* the computation of which is well known to require delicate treatment.

The lattice of models

We now investigate the structure of the space of models entailed by a given context. The resulting structure bears some similarity to that developed in "concept lattices" (see Davey & Priestley, 1990), though the motivation is different. This characterization is critical partly because it allows the full set of legal models in a given space to be enumerated. Moreover, it is essential because inference among models—e.g. the computation of the correct model for a given observed configuration—is actually carried out by movement through the internal structure of the model space. The space

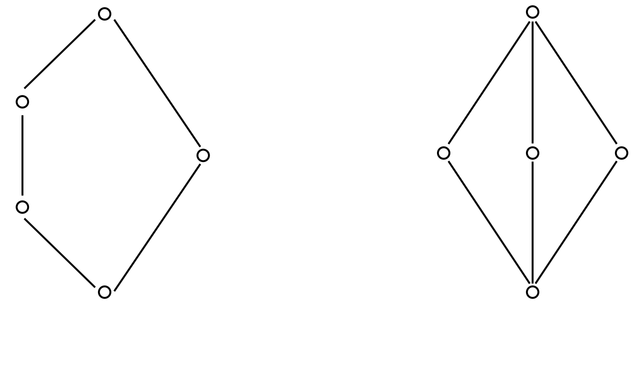


Figure 5. Two forms that do not appear in any distributive lattice.

of models turns out to be a distributive lattice (cf Erné, 1993), an extremely well-behaved type of partial order. Some properties entailed by this fact, which are extremely desirable from the point of view of perceptual inference, will be exhibited below.

Recall that a *lattice* is a partial order in which every pair of elements A and B has a unique greatest lower bound (called the "meet," $A \wedge B$) and least upper bound (called the "join," $A \vee B$. (See Grätzer, 1978 or Davey & Priestley, 1990 for good introductions to lattice theory.) Pictorially, a lattice is a partial order in which any two nodes have a unique common bound (i.e. a node they both connect to) both above and below. A *distributive lattice* is one that obeys the "distributive identities"

(8)
$$\begin{aligned} (A \wedge B) \vee (A \wedge C) &= A \wedge (B \vee C), \\ (A \vee B) \wedge (A \vee C) &= A \vee (B \wedge C), \end{aligned}$$

for all A, B, and C; that is, $\wedge$ and $\vee$ "distribute," as one would expect if they were read as intersection and union respectively. Pictorially, distributive lattices are distinguished by the fact that they contain no "pentagons" or "diamonds" (see Fig. 5) as sublattices. This rule serves as an thumbnail check that lattices expected to be distributive actually are so, and can be confirmed later in Figs. 6 and 12-15.

One might show that the set of models is a distributive lattice simply be confirming the distributivity condition algebraically. However it is far more revealing of the internal structure of models to do so by another means—namely by explicitly constructing an isomorphism between the set of models and a lattice that we know is distributive, but whose structure is more transparent: a *ring of sets*[3], which is a set of sets that is closed under both union and intersection.

We use a basic theorem about distributive lattices:

Theorem 2 (Birkhoff's Representation Theorem) *A lattice is distributive iff it is isomorphic to a ring of sets.*

[3] Also called a "lattice of sets."

Proof. See Davey and Priestley (1990) Thm.8.17, or Grätzer (1978) Thm. II.19.

A distributive lattice, in other words, is one that behaves as if $\wedge$ and $\vee$ were intersection and union respectively, and which is closed under them. What is remarkable about the fact that the lattice of completed models is distributive is that, as mentioned above, model regions (whether of completed or uncompleted models) are *not* generally closed under these operations. Nevertheless, these regions can be placed in one-to-one correspondence with a set of sets which are closed under them—namely, the uncompleted models themselves. Correspondingly, taking completed models as logical expressions, we can construct connectives exactly isomorphic to intersection and union under which the set of completed models is in fact closed. These operators, which the following theorem guarantees exist, are the meet and join of the lattice of completed models.

Theorem 3 . *The set of completed models forms a distributive lattice, when regarded as a partial order under subset inclusion on the positive part of the models.*

Proof. The set of uncompleted models (regarded as sets) is closed under union and intersection. Consider arbitrary models M and N:

> *Union.* For every $R \in M$, each S entailed by R must also be in M, because M is closed under ω; likewise for N. Hence S must also be in $M \cup N$. Hence $M \cup N$ is closed, and is a model.
>
> *Intersection.* For each $R \in M \cap N$, we know that $R \in M$ and $R \in N$. Hence the closure of $\{R\}$ under ω is also in both M and in N; therefore it is also in $M \cap N$. Therefore $M \cap N$ is closed and is a model.

Therefore, by Thm. 2 above, the set of uncompleted models form a distributive lattice when regarded as a partial order under set inclusion. Because completed models are isomorphic to uncompleted models, it follows immediately that the set of completed models form a distributive lattice, when regarded as a partial order under set inclusion on the positive part of the completed models. □.

The following isomorphism maps uncompleted models to completed models, and intersection and union to corresponding novel operators on completed models.

(9)
$$\begin{aligned} M &\iff M^* &&= \overbrace{M}^{+} \wedge \overbrace{R - M}^{\neg} \\ M \cap N &\iff M^* \vee^* N^* &&= \overbrace{M \cup N}^{+} \wedge \overbrace{R - M \cup N}^{\neg} \\ M \cup N &\iff M^* \wedge^* N^* &&= \overbrace{M \cap N}^{+} \wedge \overbrace{R - M \cap N}^{\neg}. \end{aligned}$$

Note the seemingly counterintuitive polarity of this mapping, in which $\cap$ goes to $\vee^*$ and $\cup$ to $\wedge^*$. This serves to preserve the sense of the correspondence between composition of atomic regularities and composition of models; defined this way, an expression built from $\wedge^*$, like one built from $\wedge$, is satisfied when both left and right arguments are satisfied. Interestingly, the operators can be rewritten with the more intuitive parity, but at a cost in simplicity:

(10)
$$\begin{aligned} M^* \wedge^* N^* &= \overbrace{(R - M) \cap (R - N)}^{+} \wedge \overbrace{R - (R - M) \cap (R - N)}^{\neg} \\ M^* \vee^* N^* &= \overbrace{(R - M) \cup (R - N)}^{+} \wedge \overbrace{R - (R - M) \cup (R - N)}^{\neg}. \end{aligned}$$

For any two completed models, their meet on the lattice is simply the model which obeys all regularities that either of them obeys; their join is the model which obeys only those regularities that both obey, and fails on the others. Movement down the lattice thus always yields a strictly more regular model, and movement up the lattice always yields a strictly less constrained (more generic) model. We will sometimes denote the model lattice for a context C by $L_M(C)$ (omitting the C when the choice of context is clear), and the corresponding partial order by $\overset{M}{\leq}$, i.e.

(11)
$$M_1^* \overset{M}{\leq} M_2^* \text{ iff } M_1 \supseteq M_2.$$

Because of the Genericity Constraint, the lattice partial order is actually a preference ordering: lesser models, lower on the lattice, are always preferred. Hence the partial order is instrumental in inference, as will be detailed below.

The lattice contains all the models entailed by a given context, and completely diagrams the relationships among them. Fig. 6 shows lattices for several example contexts.

First, one remark about the internal structure of these lattices is required. Recall that the codimension of a model (the number c from Eqs. 5 and 7) is the number of regularities in the (positive part of the) model, and hence is a measure of the degree of regularity of the model. Notice in Fig. 6 that each lattice is built of rows containing models of the same codimension, with the codimension-0 (completely generic) model at the top, and the highest-codimension (completely regular) model at the bottom. This is a natural reflection of the fact that the lattice partial order is an ordering on degree of regularity. However, the correspondence between codimension and row number cannot be taken for granted, because in an arbitrary lattice, the "row number" of a given model may not even be well-defined. This is because different paths from the top to the given model may not have the same length. However, a property of distributive lattices guarantees that the row number actually is well-defined, thus allowing us to think of "number of regularities" and "row number" interchangeably when referring to codimension.

For a certain fixed model in a lattice, consider the length of paths connecting it to the top node to the model. A *chain* in a partial order is a totally ordered subset. A *maximal chain* is a chain which is as "large as possible" in that for each elements A, C in the chain, if $A \leq B \leq C$, then B is also in

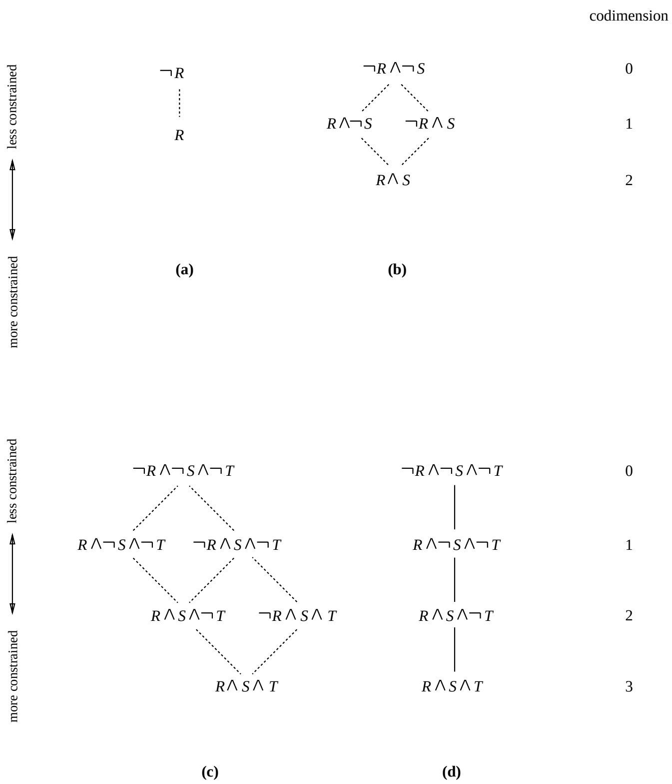


Figure 6. Model lattices for some sample contexts. **(a)** $R = \{R\}$;**(b)** $R = \{R, S\}$; **(c)** $R = \{R, S, T\}$, $T \overset{\omega}{\to} S$; **(d)** $R = \{R, S, T\}, T \overset{\omega}{\to} S, S \overset{\omega}{\to} R$.

the chain. That is, a maximal chain is a complete "path." In a distributive lattice, all maximal chains from the top node to a given node have the same length (see Grätzer (1978), Thm. IV.2.[4]) This property, called the Jordan-Hölder chain condition, can be readily confirmed by inspecting Fig. 6, and counting the length of the various paths from the top node to each fixed model. This theorem establishes that every model has a well-defined row number, which clearly corresponds[5] to codimension.

Canonically, as discussed in Section , regularity regions in fact refer not merely to subsets of a configuration space, but to regions of (one) lower dimension than the space, i.e. regions whose geometric codimension is one, this justifying the "primitive preference principle" (Eq. 1). Hence the distributivity of model lattices serves to establish not only that the discrete concept of codimension is well-defined, but also that it aligns correctly with the geometric concept.

Inference to a model. Given an observed configuration x in a context C with entailed models L_C which model should it be assigned to?

There are a number different but equivalent ways of identifying the most preferred model. Above, it was remarked that x actually satisfies some set of (uncompleted) models, but that it will satisfy exactly one of these models generically. Equivalently, x satisfies the *positive part* of some set of completed models; denote this set $\{M_x^*\}$. This set of models has a unique minimum in the partial order, which is also the model with the *maximum codimension*. This model is the unique generic interpretation of x, and is the sole completed model that x satisfies. These various equivalent definitions are captured by the following theorem, which summarizes many of the results in this section.

Theorem 4 (Equivalence of inference rules) *Given a configuration x satisfying the positive part of some set of models $\{M_x^*\}$ as described above, the following are equivalent:*

1. x is generic in model M
2. x satisfies the completed model M^*
3. $M^* = \bigwedge^* \{M_x^*\}$
4. M^* is the maximum codimension model in $\{M_x^*\}$.

We take pains to distinguish among these different but equivalent definitions of the single "best" model, because when these ideas are generalized to the grouping problem in the next section, certain of the corresponding definitions turn out *not* to be equivalent. In particular there will generally be multiple generic solutions but a unique maximum-codimension solution.

This completes the exposition of the theory of "simple" models and the inferential machinery relating to them, culminating in the statement of the maximum-codimension rule for selecting the most preferred model. The next section takes up the question of how such models can be computed concretely, which is necessary before actual computational examples can be exhibited in Section .

Regularities and models in a Logic Programming setting

Above, each regularity R has been treated as an abstract logical predicate. Naturally, we would like to evaluate these expressions computationally. This is a problem taken up by Logic Programming (see Apt (1990) for an introductory survey). One of the major issues in Logic Programming is what must be assumed in order to justify equating the truth or falsity of logical expressions with the success or failure of some corresponding computational procedure. Such assumptions touch on the "semantics" underlying automatic inference, in that they reflect how perfectly a given computational model can be regarded as modeling the world in question. We will appropriate some well-established arguments from this field in order to gain an insight onto the semantics of perceptual interpretations.

For each regularity R we assume that there is a corresponding computable clause $\mathtt{R}$ that can be carried out on a configuration x such that

(12) $$R(x) \leftarrow \mathtt{R}[x],$$

[4] In fact, semimodularity, a weaker condition that distributivity, is all that is required for this property.

[5] Here we ignore a minor technical problem involving cycles within ω, which causes no conceptual problems but would muddy the presentation.

meaning that $R(x)$ can be definitely asserted if the clause $\mathtt{R}$ halts with success on input x. Similarly, the entire regularity set $\mathcal{R}$ has associated clauses

$$\begin{aligned} R_1 &\leftarrow \mathtt{R}_1[x], \\ R_2 &\leftarrow \mathtt{R}_2[x], \\ &\vdots \\ R_k &\leftarrow \mathtt{R}_k[x]. \end{aligned} \tag{13}$$

The actual computation of models depends on being able to assert that certain regularities do *not* hold—namely, the ones in the negative part of the completed model. Hence a critical question is: under what circumstances can we positively assert $\neg R$?

A classical solution to this question is the "negation as failure rule:" assert $\neg R$ if the clause $\mathtt{R}$ halts with failure; that is,

$$\neg R(x) \leftarrow \mathtt{R}[x] \text{ halts with failure.} \tag{14}$$

In many applications, the set satisfying a clause R turns out to be recursively enumerable but not decidable; in such a case $\mathtt{R}$ cannot be assumed to halt with failure if $\neg R$, because it may not halt at all. Here, we have a different but analogous situation. We can assume that each clause will halt on every x. Nevertheless, regularities should really be regarded as referring to the world, not the observation; hence the question remains of whether a given world configuration can be "proven regular" given a particular regularity definition. If the regularity clause fails, then, what must be assumed in order to justify the inference that the world is in fact not regular in the intended sense?

Reiter (1978) observed that this inference can be regarded as depending on an assumption about the *world* under consideration, namely that it is closed—i.e. states in which R obtains but $\mathtt{R}$ fails anyway do not exist:

> **Closed World Assumption (regularities version):** For each regularity R, the corresponding computable clause $\mathtt{R}$ is complete.

This Closed World Assumption (CWA) can simply be regarded as augmenting the original clausal definition

$$R(x) \leftarrow \mathtt{R}[x], \tag{12}$$

with its converse

$$R(x) \rightarrow \mathtt{R}[x] \tag{15}$$

to yield the "completed" definition

$$R(x) \leftrightarrow \mathtt{R}[x], \tag{16}$$

from which we can conclude Eq. 14 by modus tolens (Clark, 1978). With this equivalence established, we can drop the distinction in typeface between $R(x)$ and $\mathtt{R}[x]$. By adopting the CWA, we are simply assuming that regular states of a type that our definitions cannot uncover do not happen—or at least that we will disregard cases where they do. In fact such cases, usually labeled "misses," are ubiquitous in perceptual theory, which is in effect built on the tacit assumption that ignoring them will not lead to too many incorrect conclusions. Here we are attempting to make this assumption more explicit.

The above version of the CWA allows the semantic scope of a particular regularity inference to be stated precisely. Clearly, this aids in understanding the scope of model inferences as well, but it does not completely settle the matter. We have only assumed so far that each regularity is closed. In order to justify the inference of an entire model, we need to assume further that the regularity *set* itself is closed—i.e. that no regularities other than those in $\mathcal{R}$ exist in our world.

> **Closed World Assumption (models version):** $\mathcal{R}$ is complete.

Recall that the basic model definition was "A model $M = R_1 \wedge R_2 \wedge \cdots \wedge R_c$ holds generically on x if the regularities $R_1(x) \ldots R_c(x)$ hold, but no other regularities hold". Now due to the CWA, this relatively abstract formulation can be reified computationally in the fully completed form

$$M^*(x) \leftrightarrow \begin{cases} \left.\begin{array}{c} \mathtt{R}_1[x] \\ \vdots \\ \mathtt{R}_c[x] \end{array}\right\} \text{halt with success} \\ \left.\begin{array}{c} \mathtt{R}_{c+1}[x] \\ \vdots \\ \mathtt{R}_k[x] \end{array}\right\} \text{halt with failure.} \end{cases} \tag{17}$$

This expression can be thought of as a fully computationalized version of Eq. 7, and renders the logical definition of a model fully computable. Now, by virtue of having characterized the scope of model assignment in this careful fashion, via the uniqueness of the model picked out by Thm. 4, we obtain the following characterization of the semantics of the generic model:

Theorem 5 *For a given configuration x and a regularity set $\mathcal{R}$, there is a unique computable interpretation M^* that satisfies both of*

1. the Genericity Constraint
2. the Closed World Assumption.

Proof. Because the correspondences in Eqs. 14 and 16 are both if-and-only-if, this is an immediate consequence of Thm. 1. □

This theorem summarizes the unique epistemic status of the regularity-based interpretation of the observed configuration.

We can now proceed with the main goal of the paper, the generalization of this machinery to the grouping problem. The basic theory of grouping includes suitable hierarchical counterparts of each of the key elements of the simple theory: models, codimension, the lattice of models, and so forth. Like the simple theory, the hierarchical theory will culminate in an inference rule that selects a most-preferred generic interpretation.

Grouping as regularity-finding

The Dots World

All of the above machinery has been implemented in Prolog, using an efficient approximation described in Section , which generated all of the following examples and pictures.

Consider the "Dots World." Here each configuration $\mathbf{x}$ consists of a set of n points ("dots") in the plane

(18) $$\mathbf{x} = \{x_1, x_2, \ldots, x_n\} \subset R^2.$$

The idea is to regard the process of grouping these points (dots) as a side-effect of finding the "structure" or regularity in the configuration (Witkin & Tenenbaum, 1983). The first step is to choose a regularity set R , designating which classes of configurations are to be regarded as special. In the Dots World, one obvious choice is *collinearity*, defined over dot triplets $\{x_1, x_2, x_3\}$:

(19) $$\texttt{collinear}(x_1, x_2, x_3) \leftarrow \mathrm{abs}(\pi - \backslash x_1 x_2 x_3) < \theta_{\mathrm{coll}},$$

where θ_{coll} is some threshold angle ($\pi/3$ is used in the examples later). Fig. 7 shows an example; notice the isomorphism between this figure and Fig. 6a). Collinearity is well known to be perceptually important, serving as a cue to an underlying generating curve or contour, both computationally (Parent & Zucker, 1989) and psychologically (Feldman, 1996).

Another special configuration of dots that is unlikely to occur by accident is *coincidence*:

(20) $$\texttt{coincident}(x_1, x_2) \leftarrow \|x_1 - x_2\| < \theta_{\mathrm{coinc}},$$

where θ_{coinc} is some threshold distance. This succeeds when x_1 and x_2 fall near each other. Like collinearity, coincidence is well known to be employed by human observers. Adopting it as a regularity is obviously related to the Gestalt principle of proximity. But notice that the reasoning here is actually reversed from the usual. Instead of a somewhat arbitrary presumption about inference ("if two objects are near each other, group them together"), we have one about what structures occur in the world ("two independent objects are unlikely to fall near each other by accident"). This assumption impacts on inference only via the need for a generic interpretation, in which coincident items must be interpreted as being non-independent. Rather than presuming a reasoning principle, here we assume only that independent but nevertheless coincident items—like non-generic viewpoints, winning lottery tickets, and other low-probability events—are in fact atypical events in the world.

Parse trees. It is convenient to notate the regularity $\texttt{coincident}(x_1, x_2)$ as a *tree*:

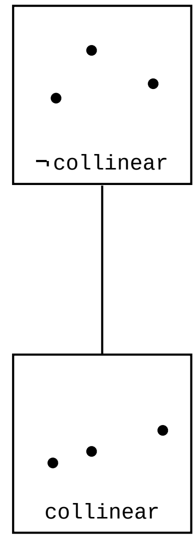
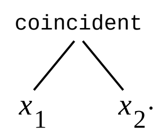


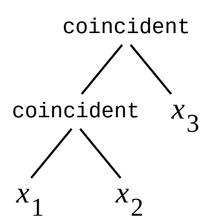
Figure 7. A dot triplet satisfying (bottom) and not satisfying (top) the regularity `collinear`. Note the isomorphism between this figure and Fig. 6a.

(21)



Each leaf in such a tree may be filled by any object that has a location, not just a simple dot. For example, the above tree itself has a location (e.g. the centroid of the x_1 and x_2). Placing a tree in one of the leaf slots yields a more complex tree that describes a coincidence among three dots:

(22)



This process can be carried out ad infinitum, yielding trees of arbitrary depth. We call these trees *parse trees*, following the linguistic parsing literature. Parse trees can be thought of as descriptions or models of arbitrary large sets of dots.

In light of this construction, it is simplest to rewrite the `collinear` regularity, which above was written as a regularity among three dots, as a binary relation between a tree and a dot.

(23) $\texttt{collinear}(T, x) \leftarrow x$ falls on the line defined by T,

where T is a parse tree that passes up an orientation and a location (see below); this definition is equivalent to Eq. 19. We assume, following apparent human intuitions, that the two arguments of the redefined `collinear` can only be judged collinear if they are also near enough to be judged `coincident`:

(24) $$\texttt{collinear} \xrightarrow{\omega} \texttt{coincident},$$

i.e. a collinearity among three dots occurs when two dots are near each other (satisfy `coincident`), a third dot is near the first two, and the third dot falls on the line defined by the first two.

Summarizing, we have the following context for the Dots World:

(25)
$$R \;:\; \{\texttt{coincident}, \texttt{collinear}\},$$
$$\omega \;:\; \texttt{coincident} \xrightarrow{\omega} \texttt{collinear},$$

which yields the model lattice

(26)
$\emptyset$
|
{coincident}
|
{collinear, coincident} .

For simplicity, we abbreviate this to

(27)
gen
|
coincident
|
collinear .

keeping in mind that, strictly, `gen` means $\emptyset$, `coincident` means $\{\texttt{coincident}\}$, and `collinear` means $\{\texttt{collinear}, \texttt{coincident}\}$ (i.e. the $\wedge$-expression $\texttt{collinear} \wedge \texttt{coincident}$). Recall that because of the CWA, in the chosen context `gen` really means $\neg\texttt{coincident} \wedge \neg\texttt{collinear}$.

The next section develops parse trees more completely. Each nonterminal node of a parse tree will always be one of the above three models, with the leaves being simple dots. A parse tree is thus a hierarchical version of a "model," i.e. it is an interpretation of a configuration of dots; but it is also *built* from simple models, and the formal properties of parse trees will derive in part from the formal properties of models, which is why those properties were derived in such detail in the previous section. In particular some theorems about the internal structure of the space of parse trees depend directly on facts about the structure of the lattice of models.

The grouping interpretation resulting from this machinery, the maximum-codimension parse tree or qualitative parse, is the most regular interpretation of the dot configuration (modulo the chosen regularity set): hence the phrase "grouping as regularity-finding." Coupling this with "regularity-finding as negation" (because the Genericity Constraint requires that a regular interpretation only holds if all more regular interpretations do not hold) and "negation as failure" (because of the CWA) yields the handy motto "grouping as failure."

The qualitative parse

We now present the grouping theory more formally, following the outline of the simple theory. Each of the following definitions is simply a recursive version of an idea presented above. We begin with a recursive definition of a parse tree. Trees are notated in pseudo-Prolog as lists. In each tree other than `[dot]`, the head term (root node) is a model, and the second term is a list of arguments to the head term.

Defn. 5 (Parse tree)

1. $[\texttt{dot}]$ is a parse tree;
2. $[M, [T_1, T_2, \ldots]]$ is a parse tree if
1. $M \in L_C$ (i.e. M is a model), and
2. $T_1, T_2, \ldots$ are parse trees.

Parse trees can be satisfied by a set of dots coupled with an *assignment* of those dots to the leaves of the tree. A complete assignment in a complete parse tree is conveniently notated by substituting the name of the dot in place of the tree $[\texttt{dot}]$ in the appropriate leaf. A parse tree coupled with an assignment is called an *interpretation.*

The tree $[\texttt{dot}]$ is satisfied by any dot. Satisfaction of other trees is mediated by parameter passing. Each regularity, when it succeeds, passes appropriate parameters up the parse tree to be fed as arguments to regularities on the next higher level. $[\texttt{dot}]$ passes the coordinates of the dot. `coincident` passes the centroid of the coincident dots, which is then treated exactly as if it were dot. `collinear` passes the coordinates of the last two dots in the collinear chain, which define an orientation. Then, satisfaction of a tree and generic satisfaction of a tree are both defined in a straightforward recursive manner: a tree is satisfied (generically) if its head term is satisfied (generically) and each of its subtrees is satisfied (generically).

Defn. 6 (Satisfaction and generic satisfaction of a parse tree)

1. $[\texttt{dot}]$ holds (generically) on any dot;
2. $[M, [T_1, T_2, \ldots]]$ holds (generically) on a set of dots $\{x_1, x_2, \ldots\}$ if there exists a partition P of into cells $p_1, p_2, \ldots$, such that:
1. each tree T_i holds (generically) on the dots p_i, and
2. the model M holds (generically) on the parameters passed by $T_1, T_2, \ldots$.

The recursion always bottoms out at the leaf nodes with $[\texttt{dot}]$. In actual computation, genericity is determined by regarding head terms as completed models, and then simply computing satisfaction of the completed model. An assignment is produced by recursively partitioning a dot set. In the

Dots World regularity set, both regularities take two arguments, so these partitions are always 2-way.

The notion of codimension also has a straightforward recursive counterpart:

Defn. 7 (Codimension of a parse tree)

1. The codimension of $[\texttt{dot}]$ is 0;
2. The codimension of $[M, [T_1, T_2, \ldots]]$ is the sum of the codimension of the head term M and the codimensions of each of $T_1, T_2, \ldots$.

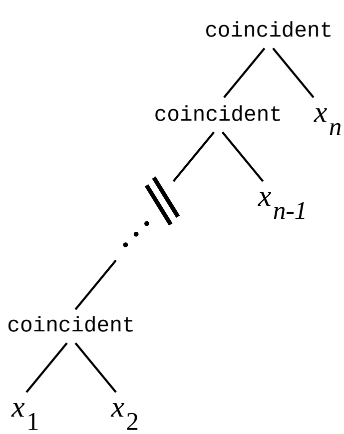
Head terms follow codim $M^* = |M|$, so codim gen = 0, codim coincident = 1, codim collinear = 2 ($= |\{\texttt{collinear}, \texttt{coincident}\}|$).[6] As mentioned above, and discussed in detail below, the generic parse tree for a given dot configuration is not unique. The most preferred interpretation, rather, is that picked out by the maximum-codimension rule:

> **Maximum-codimension rule for choosing a parse tree:** Among parse trees T in which a configuration $\mathbf{x}$ is generic, choose T that maximizes codim T.

This choice is not guaranteed to be unique, but it frequently is unique, for interesting reasons having to do with the internal structure of the partial order. The maximum-codimension interpretation will be referred to as the *qualitative parse.*

Examples. Figs 8-10 show examples of dot configurations along with their qualitative parses. The reader can perform a bit of "instant psychophysics" by evaluating whether these interpretations seem intuitively plausible, while keeping in mind that the parses can only be as psychologically compelling as the regularity sets we have defined. For the examples, θ_{coll} is set at $\pi/6$ and θ_{coinc} is set at 40 pixels (the appropriate value depends on the density of dots in the image) in a square window with sides of 360 pixels. Each figure shows the dot configuration itself (part (a) of each figure), a pictorial schematic depiction of the tree superimposed on the dot configuration (part (b)), and an explicit diagram of the tree (part (c)). The left-to-right ordering of each tree is of course meaningless. For simplicity, trees describing large chains of dots are abbreviated, so that

(28)



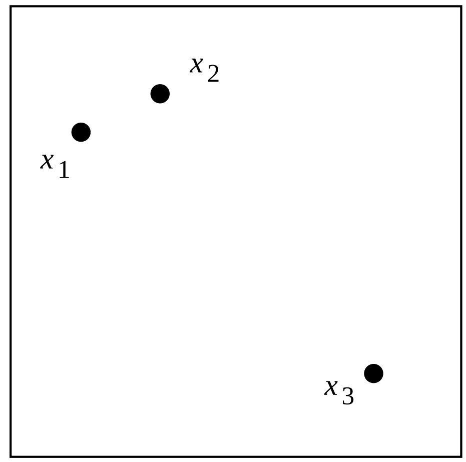
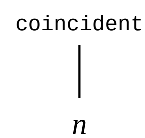


Figure 11. An example of a configuration that has two distinct generic interpretations.

is rewritten as

(29)

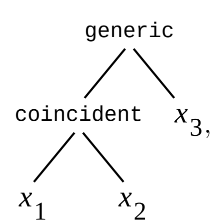


The pictorial regularity notation is as follows:

1. Dots that are coincident are drawn connected to a small hollow circle at their centroid.
2. Dots x_1, x_2, x_3 that are collinear are shown connected by line segments.
3. Dots that are mutually generic are not marked in any way, suggesting "no regular relationship." Any dot "sitting off by itself" thus explicitly indicates the failure of all relevant regularities, with respect to the other dots and groups in the field.

As mentioned above, some dot configurations have more than one generic parse tree. Fig. 11, for example, is generic under the interpretation

(30)



[6] Because coincident actually entails two fixed degrees of freedom, it might be preferable to instead choose the less strictly correct codim coincident = 2, codim collinear = 3; then the overall codimension of a parse tree would correspond to the total number of degrees of freedom generically fixed in the configuration. In fact, the two definitions agree if coincidence to regarded as two separate regularities, coincidence in the *x*-coordinate and coincidence in the *y*-coordinate, which might make sense in some contexts. The ordering properties of the resulting space of parse trees is unaffected by which definition is chosen, so we stay with the strict one.

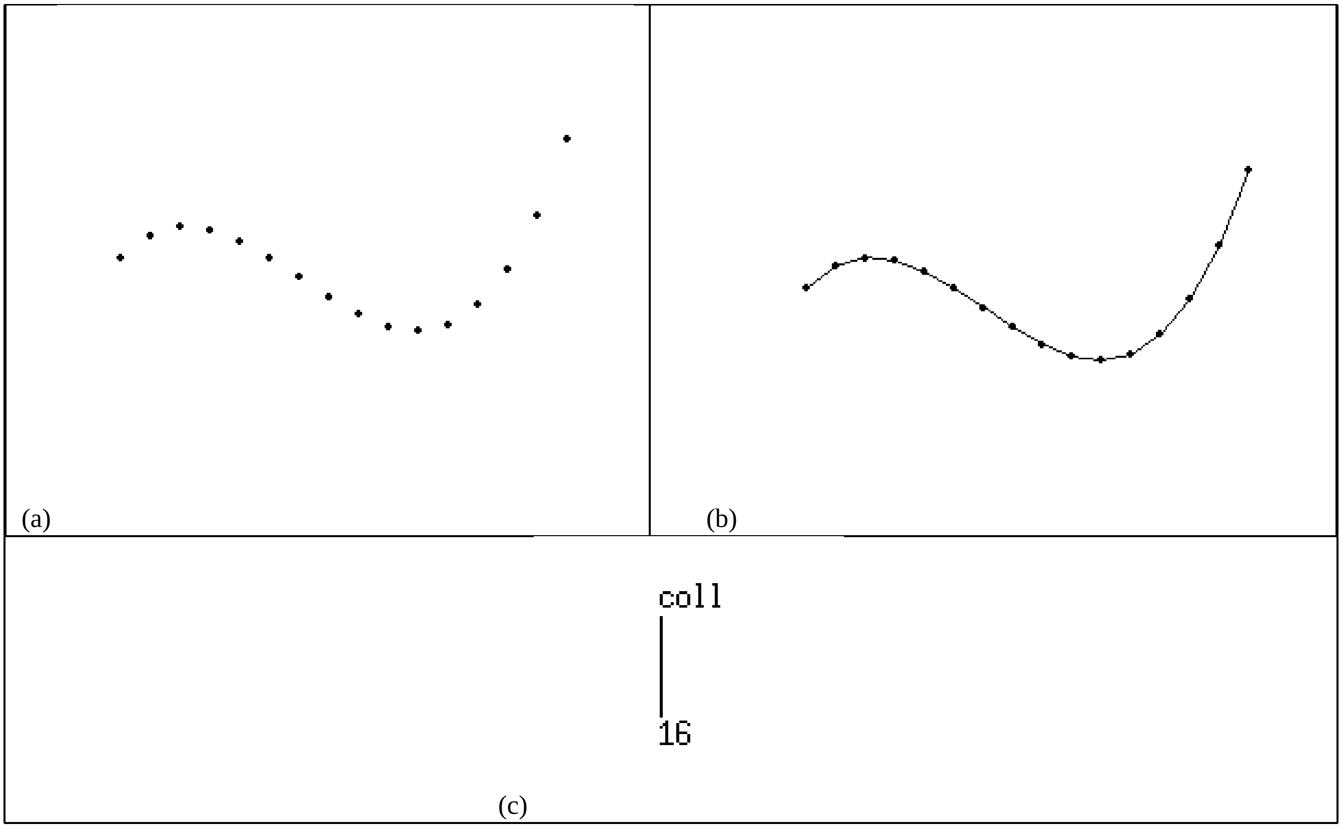
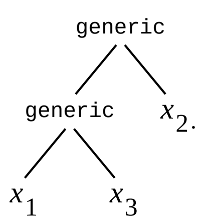


Figure 8. An example of a dot configuration and its maximum-codimension parse tree, showing **(a)** the dot configuration, **(b)** a schematic pictorial depiction of the tree superimposed on the original configuration (see text for description) and **(c)** the parse tree in explicit form.

but is also generic under the interpretation

(31)



i.e. the completely generic tree. However, tree (30) has codimension 2, while tree (31) has codimension 0. Hence the more natural interpretation, (30) and its the associated partition, is preferred under the maximum-codimension rule. The difference originates in to the top-level partition, $\{(x_1,x_2),(x_3)\}$ vs. $\{(x_1,x_3),(x_2)\}$; the former is susceptible to a more regular interpretation and is thus preferred.

The space of qualitative parses

As the lattice of models defined a partial order among models, we can now construct a partial order among parse trees, ranking them in order of degree of regularity/genericity. Like the lattice of models, this partial order can be used to completely enumerate the parse trees that are possible for a given number of dots *n*, from the most generic and unstructured interpretation to the most regular and constrained. Each of these parse trees is a possible grouping interpretation for some dot configuration. Hence the enumeration will generate every distinct grouping interpretation that can be drawn for *n* dots—an exhaustive qualitative map of the Dots World.

The structure of the space.

The partial order among parse trees will be denoted $\overset{T}{\leq}$, and the corresponding poset (the set of parse trees regarded as a partial order) by $L_T(C)$. It has a simple recursive definition building on the partial order L_M among simple models:

Defn. 8 (Partial order among parse trees) *For parse trees* $T_1 = [M_1,[T_{1a},T_{1b},\ldots]]$, $T_2 = [M_2,[T_{2a},T_{2b},\ldots]]$ *we have* $T_1 \overset{T}{\leq} T_2$ *if*

1. $M_1 \overset{M}{\leq} M_2$, and
2. $T_{1i} \overset{T}{\leq} T_{2i}$ for $i = a,b,\ldots$.

T_1 is minimally more regular than T_2 if T_1's subtrees are exactly the same as T_2's subtrees, but T_1's head term is minimally more regular than T_2's head term; or, if T_1 and T_2 have the same head term, but exactly one of T_1's subtrees is minimally more regular than one of T_2's subtrees, and the rest are

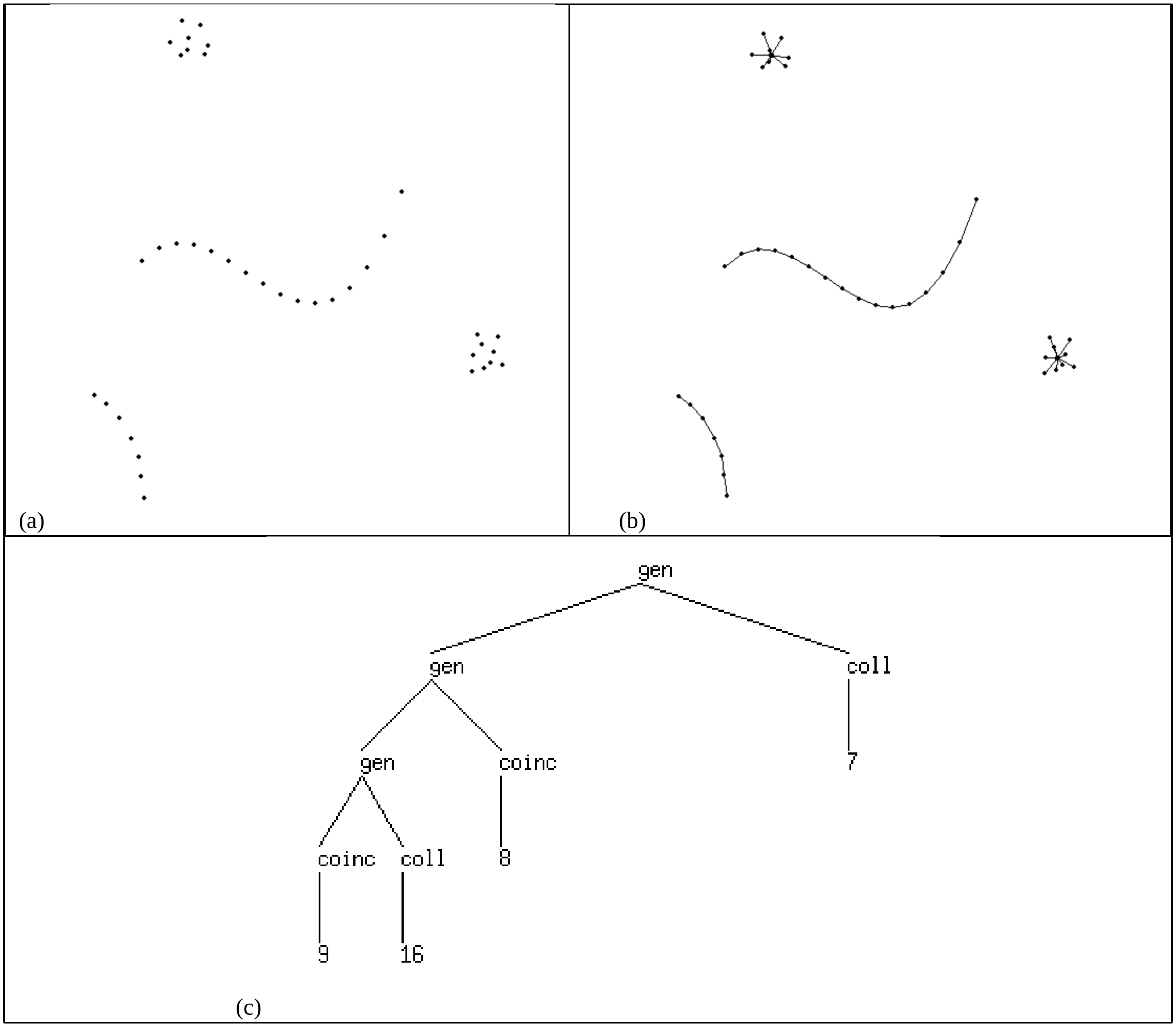
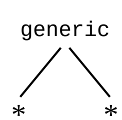


Figure 9. Another dot configuration, with several groups of different types. The parse tree accurately reflects the intuitive structure of the configuration.

the same. That is, a given tree gets more regular when either its head term or one of its subtrees gets more regular.

For a given number of dots n, the full space of parse trees can be generated by starting with a completely generic tree G, and taking the "closure under going down" in the partial order, called the *down-set* of G and conventionally denoted $\downarrow G$. (I.e. for any T, $\downarrow T = \{t | t \overset{T}{\leq} T\}$.) For $n = 2$, for example, the completely generic tree (denoting the leaves simply by an asterisk *) is

(32)



The down-set of this tree under $\overset{T}{\leq}$ is the complete space of qualitative parses for $n = 2$, and is shown in Fig. 12.

Similarly, for $n = 3$ the generic tree is

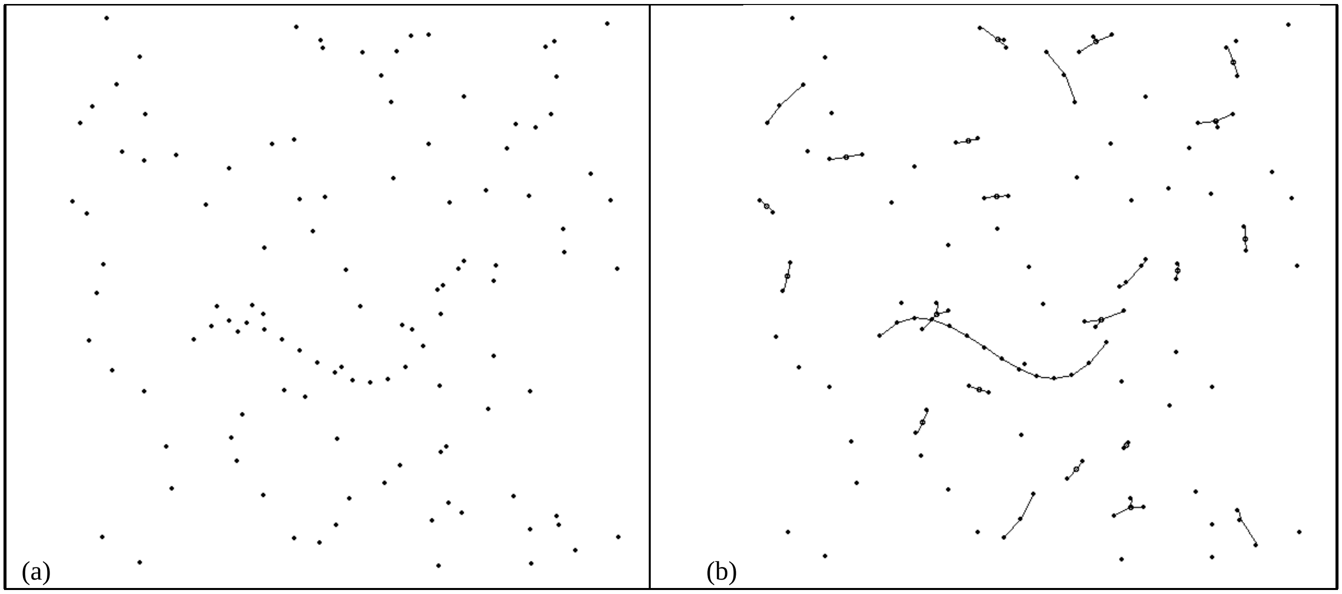


Figure 10. An example with a curvilinear chain embedded in 100 random points, which the program identifies as a high-codimension subtree. Here the full parse tree is too large to display conveniently.

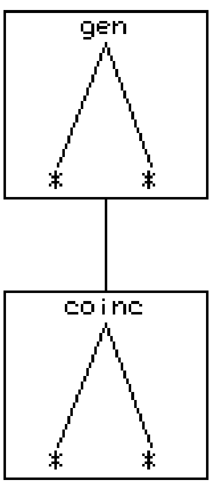
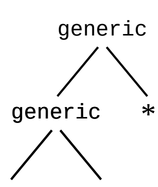


Figure 12. The space of qualitative parses for $n = 2$. The notation is the same as the individual parse trees exhibited above, except with * substituted for `[dot]` in the leaves, and `generic`, `coincident` and `collinear` abbreviated to `gen`, `coinc`, and `coll`, respectively.

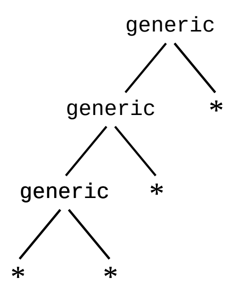
(33)



and its down-set is the space of parses for three dots (Fig. 13), which in this case is a total ordering of four parse trees. At the top is the parse tree for three mutually generic dots, a totally unstructured configuration; at the bottom is the tree for a chain of three collinear dots, a maximally structured configuration.

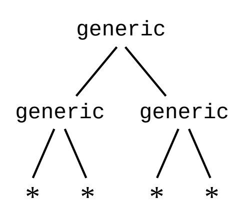
For $n > 3$ the structure gets slightly more complicated. For $n = 4$ there are *two* distinct completely generic trees,

(34)



and

(35)



The complete space of parses consists of the union of the down-sets of these trees, and is shown in Fig. 14. This partial order can be plainly seen to consist of two disjoint lattices (one of them being a total order), one hanging off each generic tree. Both of these lattices are distributive, as can

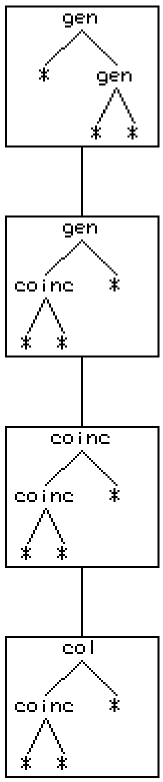


Figure 13. The space of qualitative parses for $n = 3$.

be readily confirmed using the "no diamonds or pentagons" rule. The lattice for five dots has a similar form, and is shown in Fig. 15.

In fact, this turns out to be the general form of L_T for a given number of dots n: multiple disjoint distributive lattices, each hanging off a distinct completely generic tree.

Theorem 6 *Let $G_1, G_2, \ldots$ denote the distinct completely generic trees for n, and again let $\downarrow T$ denote the down-set of a parse tree T under $\overset{T}{\leq}$. Then:*

1. Every legal parse tree is contained in some $\downarrow G_i$;
2. The $\downarrow G_i$'s are disjoint;
3. For each G_i, $\downarrow G_i$ is a distributive lattice.

Proof.

1. This is evident because for any M, $M \overset{M}{\leq}$ `generic`.

2. All the G_i's have distinct topologies, by definition. Every parse tree within a given $\downarrow G_i$ has the same topology, namely that of G_i. Hence trees in different $\downarrow G_i$'s must have different topologies.

3. By induction on n. The direct product of distributive lattices is distributive (Davey & Priestley, 1990, Proposition 6.8). In general,

$$\downarrow [M, [T_1, T_2, \ldots]] \cong \downarrow M \times \downarrow T_1 \times \downarrow T_2 \ldots,$$

where $\cong$ indicates an isomorphism. Each G_i on n dots is really $[\texttt{generic}, [g_1, g_2, \ldots]]$, where $g_1, g_2, \ldots$ are each generic trees on some $m < n$, so

$$\downarrow G_i \cong \downarrow \texttt{generic} \times \downarrow g_1 \times \downarrow g_2 \ldots .$$

Now, $\downarrow$`generic` is just $L_M(C)$, which is distributive by Thm. 2. Therefore $L_T(C)$ for n is distributive if $L_T(C)$ for $m < n$ is distributive. $L_T(C)$ for $n = 2, 3, 4, 5$, exhibited above[7] are distributive (by inspection); hence so is $L_T(C)$ for arbitrary n. □

Note that the distributivity of these lattices is a direct result of the distributivity of the model lattices; the partial order on head terms (i.e., models) is itself embedded in the partial order on trees, because one way for a tree to get more regular is for its head term to get more regular.

It is intriguing to regard each qualitative parse as a distinct way that the regularity set can *fail*, and the space of parses as comprising the legal "patterns of failure." Normally, of course, each regularity succeeds or fails depending on its arguments. But it is a simple matter to jury-rig each regularity so that it succeeds or fails regardless of input, making the choice non-deterministically. Then, each solution that Prolog finds on backtracking corresponds to a distinct alternative parse tree. This procedural approach is equivalent to the logical definition given above, but serves to emphasize the link between negation (failure) and the completeness of the space of interpretations.

Uniqueness of the qualitative parse.

The qualitative parse is not unique in general. However, in the Dots World context, multiple maximal solutions turn out to be unusual, in that they only occur under somewhat rarefied formal circumstances. When they do occur, they correspond to perceptually ambiguous configurations exactly as one would expect. This section will characterize the special circumstances that must occur in order to bring this about.

First of all, for each dot configuration **x** there is exactly one generic solution per disjoint lattice in $L_T(C)$. To see why, consider two parse trees T_1 and T_2, both of which hold on some **x**. If they are on the same lattice, then neither one is generic on **x**, because their meet $T_1 \wedge T_2$ exists and also holds on **x**. Hence at most one parse tree per lattice can hold generically on **x**. On the other hand, at least one parse tree on each lattice will hold on **x**, namely the generic tree at the top. Consequently, **x** has exactly one generic interpretation on each of the disjoint lattices, namely the meet of all the parse trees on that lattice that **x** satisfies.

Dot configurations, then, can have multiple generic interpretations. The question of uniqueness hangs on whether it is possible for two of them to have the same codimension—pictorially, for the best parse tree on two disjoint lattices to

[7] There is no space for $n = 1$ because no regularities in the context take only one argument.

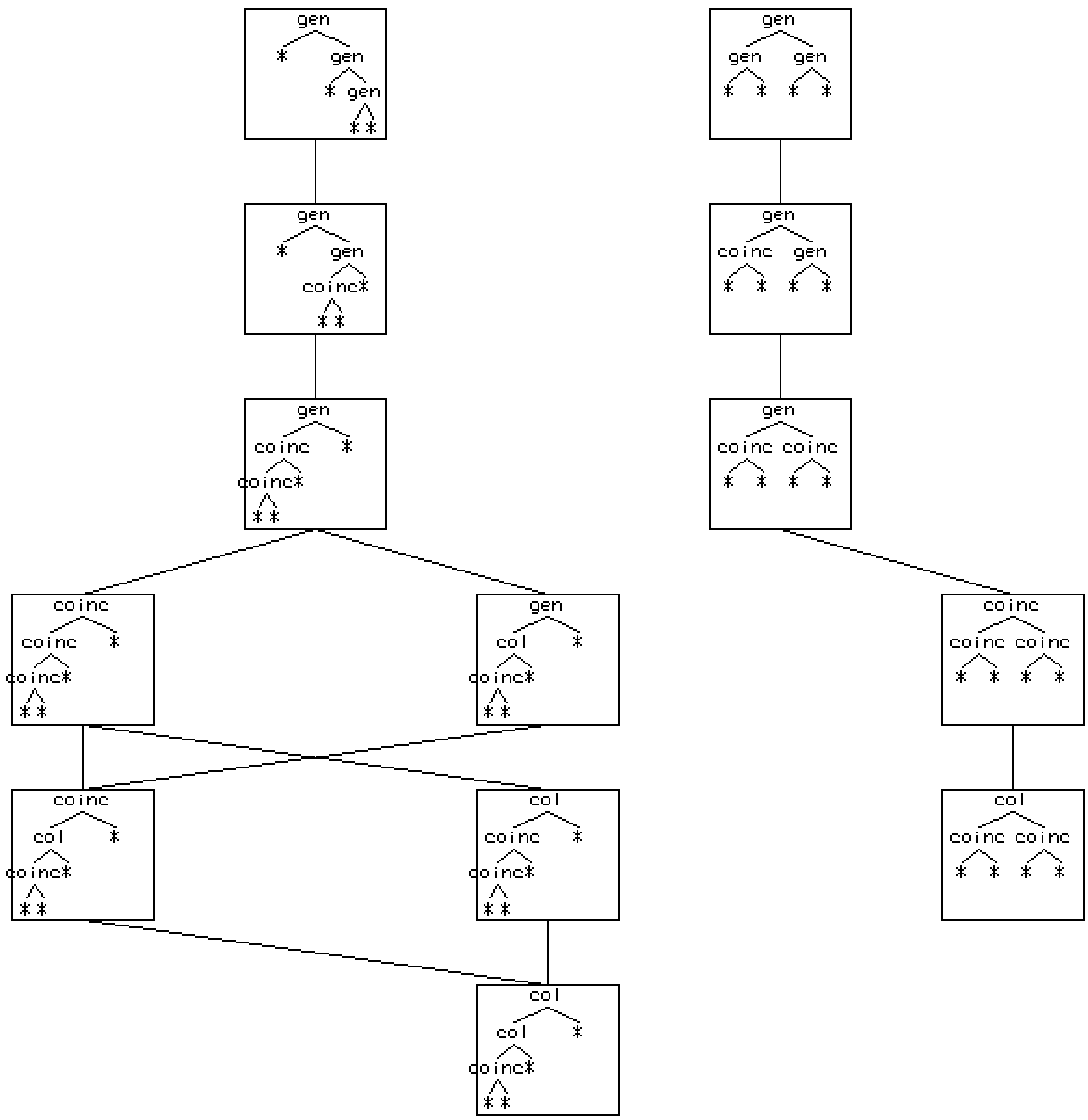


Figure 14. The space of qualitative parses for $n = 4$.

be at the same row. It turns out that the multiple generic interpretations are usually "tantamount to the same interpretation," i.e they are actually equivalent to one another in a precise sense defined below. Moreover, it turns out that under a general assumption about the context (satisfied by the Dots World), even non-equivalent same-codimension solutions usually cannot be satisfied generically by the same configuration, meaning that configurations usually have unique maximum-codimension interpretations.

Cluster-equivalence. First, we need to define what we mean by "tantamount to the same interpretation." Consider again the trees (34) and (35). They are distinct trees and cannot be brought into isomorphism with each other. Nevertheless, there is clearly a sense in which these are the same interpretation: both describe a "cluster" of four mutually generic dots. Similarly, *n* mutually `coincident` or mutually `collinear` dots can be written in a number of distinct ways. The difference between them is simply an artifact of the fact that each tree is binary. The following definition captures the necessary equivalence, which we call *cluster-equivalence*.

First, some notation: for any tree T, denote its head term (model) by $h(T)$. Denote by $m(T)$ the multiset[8] of head terms in all of T's subtrees, or recursively in their subtrees etc. (i.e. $m(T)$ includes all the models in the scope of T).

[8] i.e., set with duplicates.

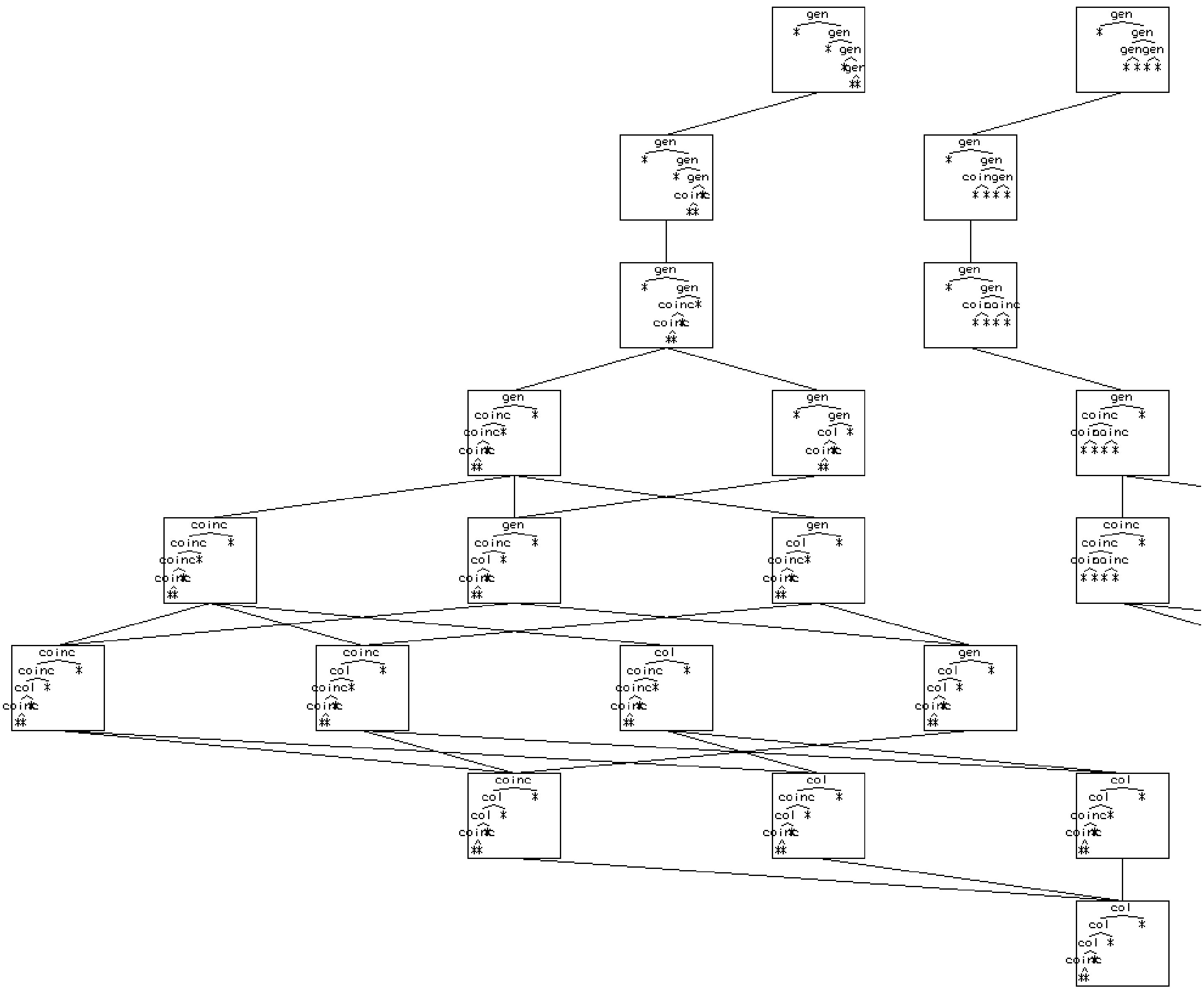


Figure 15. The space of qualitative parses for $n = 5$.

Similarly, given an assignment of dots, denote the set of dots assigned to leaves of T by $d(T)$.

Defn. 9 (Cluster-equivalence) *For parse trees* $T_1 = [M_1, [T_{1a}, T_{1b}, \ldots]]$, $T_2 = [M_2, [T_{2b}, T_{2b}, \ldots]]$, *we have* $T_1 \overset{c}{\sim} T_2$ *if*

1. the members of $m(T_1)$ (i.e., the head terms in the scope of T_1) are all the same, and $m(T_1) = m(T_2)$.
2. $M_1 = M_2$, and $T_{1i} \overset{c}{\sim} T_{2i}$ for $i = a, b \ldots$.

Cluster-equivalence is an equivalence relation, and is preserved under change in the order of subtrees, which we have been disregarding anyway. Trees (34) and (35) are a minimal case of non-isomorphic but nevertheless cluster-equivalent parse trees. Cluster-equivalent pairs bear a syntactic distinction which, like differences in the order of a tree's subtrees, does not correspond to any possible semantic distinction.

Now, we would like to show that for a configuration $\mathbf{x}$, its maximum-codimension parse tree T is unique up to cluster-equivalence. We break this proposition down into three cases, isolating the one case (Case 3 below) in which uniqueness fails. An important role is played by *transitive* regularities, i.e. regularities R such that

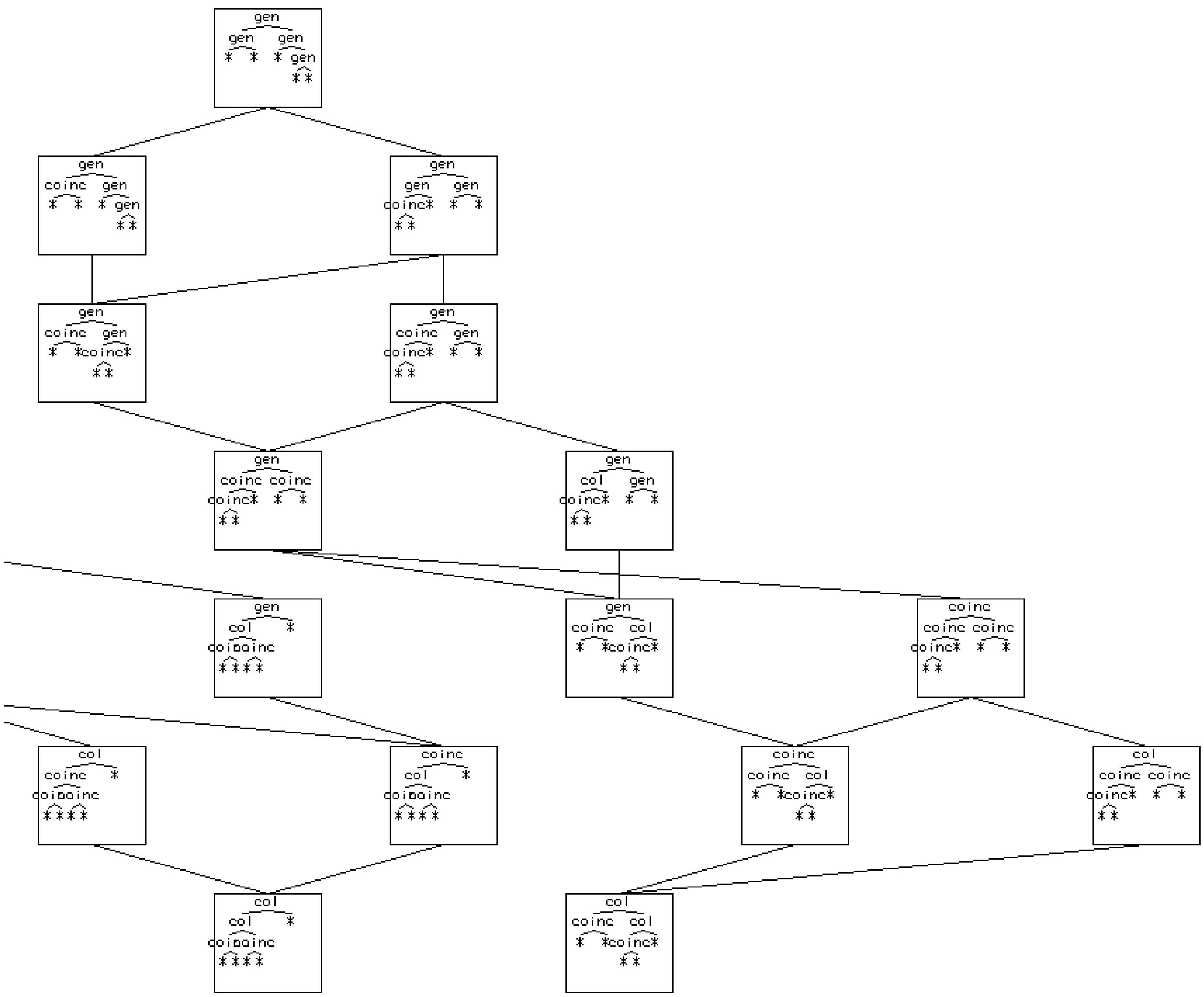
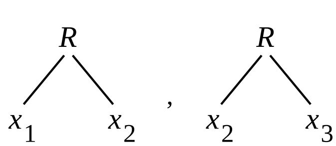
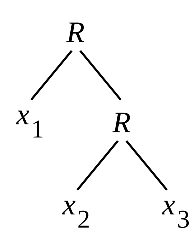


Figure 15. (continued)

(36) 

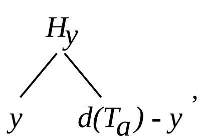
implies

(37) 

In particular, in some of the cases below we will assume that the context C has a *transitive universal consequent*, i.e. a transitive regularity R_0 such that for each $R \in \mathcal{R}$ other than

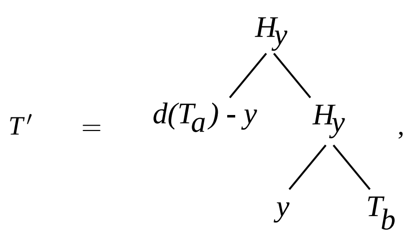
R_0, $R \xrightarrow{\omega} R_0$. In the Dots World (under a slight modification of the regularity definitions given above), `coincident` fills this role.

We start by assuming that the configuration $\mathbf{x}$ has a maximum-codimension interpretation $T = [M, [T_a, T_b]]$ (the argument would apply equally to trees with more than two subtrees). We then pursue a pigeonhole strategy, turning on the fact that in any alternative solution maximum-codimension interpretation $T' = [M', [T'_a, T'_b]]$, the subtree T'_a must span some dots from T_a's scope as well as some from T_b's scope—that is, some of the dots "cross scopes" in constructing T'. Denote by y the set of dots from $d(T_a)$ that are also in $d(T'_a)$. Denote by H_y the generic model of the relationship between y and the rest of $d(T_a)$, that is, the model such that

(38) 

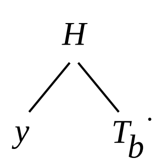
holds generically. (Here we abuse the notation slightly and let a set of dots stand in for a tree describing them.) We break the question down into three cases according to the value of M. The arguments pertaining to each case are sketched informally.

Case 1: $M = H_y$. Because y has the same relationship to the rest of T_a as T_a has to T_b, it is clear that y can be "re-attached" to T_b instead of T_a without any loss in codimension. Hence an alternative overall interpretation T' can be written as

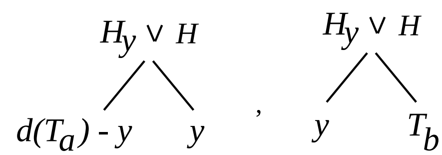
(39) 

which is cluster-equivalent to T. Hence no genuinely distinct solutions occur in this case. In practice this is the most common case.

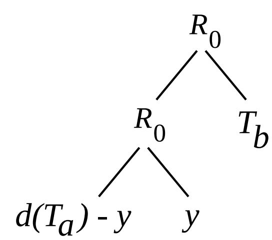
Case 2: $M =$ `generic`. This is the most interesting case; here the transitivity of R_0 actually rules out alternative interpretations. Denote by H the generic model of the relationship between y and T_b, i.e.

(40) 

Then the joined model $H_y \vee H$ holds over both the trees

(41) 

Now, if $R_0 \notin H_y \vee H$, then $H_y \vee H =$ `generic`, in which case T' cannot have the same codimension as T, contrary to assumption. So $R_0 \in H_y \vee H$, which means that by transitivity of R_0,

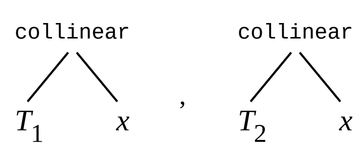
(42) 

must hold. But this means that M cannot be generic, contrary to assumption.

In other words, T must be unique because any alternative solution would have to posit some relationship between dots in its left subtree and dots in its right subtree; but if such a relationship holds, then part of it must have a transitive component (because there exists a transitive universal consequent), which means that T's head model is not generic. For example, two clusters of `coincident` dots that are generic with respect to each other must have a unique maximum-codimension parse tree; because any alternative tree would have to include some relationship between the two clusters, which if it existed would mean that the two clusters were not really generic with respect to each other.

Case 3: In the remaining cases, the transitive component of $H_y \vee H$ is also recognized by M, so there is no contradiction. The remaining component is intransitive, so the fact that it holds between the two pairs (Eq. 41) does not entail anything about M. Hence the intransitivity allows for the genuine possibility of two distinct maximum-codimension parses.

In the Dots World, `collinear` is intransitive, and indeed it is possible to construct configurations with two alternative qualitative parses that hinge on the intransitivity. Fig. 16 shows an example. There, we have

(43) 

but not

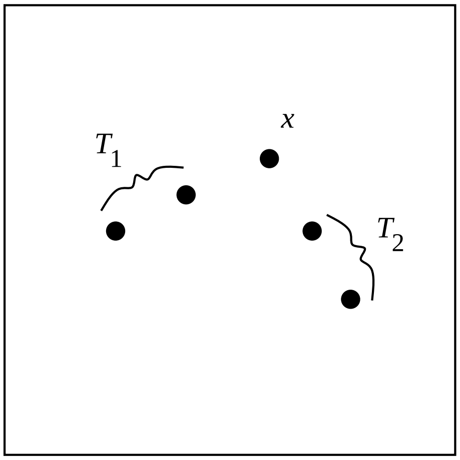
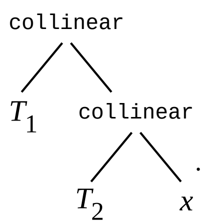


Figure 16. An example of a configuration that has two distinct maximum-codimension interpretations. Does x belong with T_1 or T_2?

(44)



The dot x can be passed between the two trees without changing codimension, resulting in two interpretations that are both generic but are not cluster-equivalent. Notice that such a configuration is itself atypical (meaning literally that it has positive codimension). Finally notice that the its two interpretations correspond neatly to a perceptual ambiguity: x shifts from grouping with T_1 to grouping with T_2. (Of course, our assignment definition does not allow x to be assigned to both trees, which might correspond more closely to the human interpretation.)

Summarizing, the qualitative parse is always locally unique (that is, it is a local minimum in the partial order $L_T(C)$), and it is globally unique in all but atypical circumstances. Coupling this fact with the semantic uniqueness of the generic (simple) model afforded by the CWA yields a similar statement of semantic uniqueness for grouping interpretations:

Theorem 7 *For a given dot configuration* **x***, the qualitative parse is the only computable interpretation that*

1. is a global maximum of regularity (i.e. has maximum codimension), and
2. satisfies the Closed World Assumption.

Like its counterpart for simple models, this statement articulates the unique epistemic status of the qualitative parse. Given the choice of description language (i.e., the regularity set), preference for any other interpretation would mean that the observer was accepting a needlessly large number of unexplained coincidences in the data.

The collapsed space. Distinct but cluster-equivalent parse trees are semantically equivalent, and it follows that they ought to be psychologically equivalent as well. It should be possible, then, to exhibit the complete set of psychologically distinct grouping interpretations, simply by collapsing each equivalence class of parse trees into a single tree T (chosen arbitrarily) which represents the class. For example, the set of distinct completely generic trees would collapse to a single generic tree, representing "the" generic interpretation.

This new smaller set of parse trees, in which each tree represents an equivalence class of cluster-equivalent trees, can then be placed into a partial order in a natural way: $T_1 \overset{T}{\leq} T_2$ if there exist some T_1 in T_1's class and T_2 in T_2's class such that $T_1 \overset{T}{\leq} T_2$. The resulting partial order is a "join semilattice" (join always exists but meet only sometimes exists). Hence it can have multiple minima. Wherever meet exists, the distributive equalities hold. Hence, critically, the Jordan-Hölder condition holds and codimension is well-defined. In effect this is because cluster-equivalent trees always have the same codimension, so codimension is preserved in the collapsed partial order. An example of this collapsed space, that for $n = 4$, is shown in Fig. 17.

The collapsed partial order is in a sense the most meaningful enumeration of grouping interpretations and their interrelationships. In practice, though, computation must be performed in the original space, as the partial order can only be cleanly defined among the raw lattices.

Discussion

The qualitative parse, like human perceptual interpretations, is very robust under slight changes in the configuration. This is inherent in its qualitative definition, which collapses over any changes that preserve regularity satisfaction. If the regularity set is well chosen, the result is a representation that is expressive, flexible, and captures stable world properties generically.

Moreover, the regularity machinery can be easily adapted simply by changing the regularity set. One obvious mild alteration would be the addition of a `smooth` regularity to supplement `collinear`, in effect adding an additional degree of continuity of the derivative of the underlying dot-generating curve. This would allow the detection of "corners" in chains of collinear dots (Link & Zucker, 1987) each of which would show up as a ¬`smooth` head term over two `smooth` subtrees. Another psychologically plausible regularity is *equal spacing* of dots along the generating curve. Feldman (1997a) reports evidence that human subjects treat equal spacing (like collinearity) as evidence in favor of a curvilinear interpretation for dot chains—that is, they treat it as a regularity. Yet another improvement would be to change the threshold definition of `coincident` to something more psychologically plausible; most evidence points to something closer to an exponential decay (Zucker, Stevens, & Sander, 1983; Barchilon

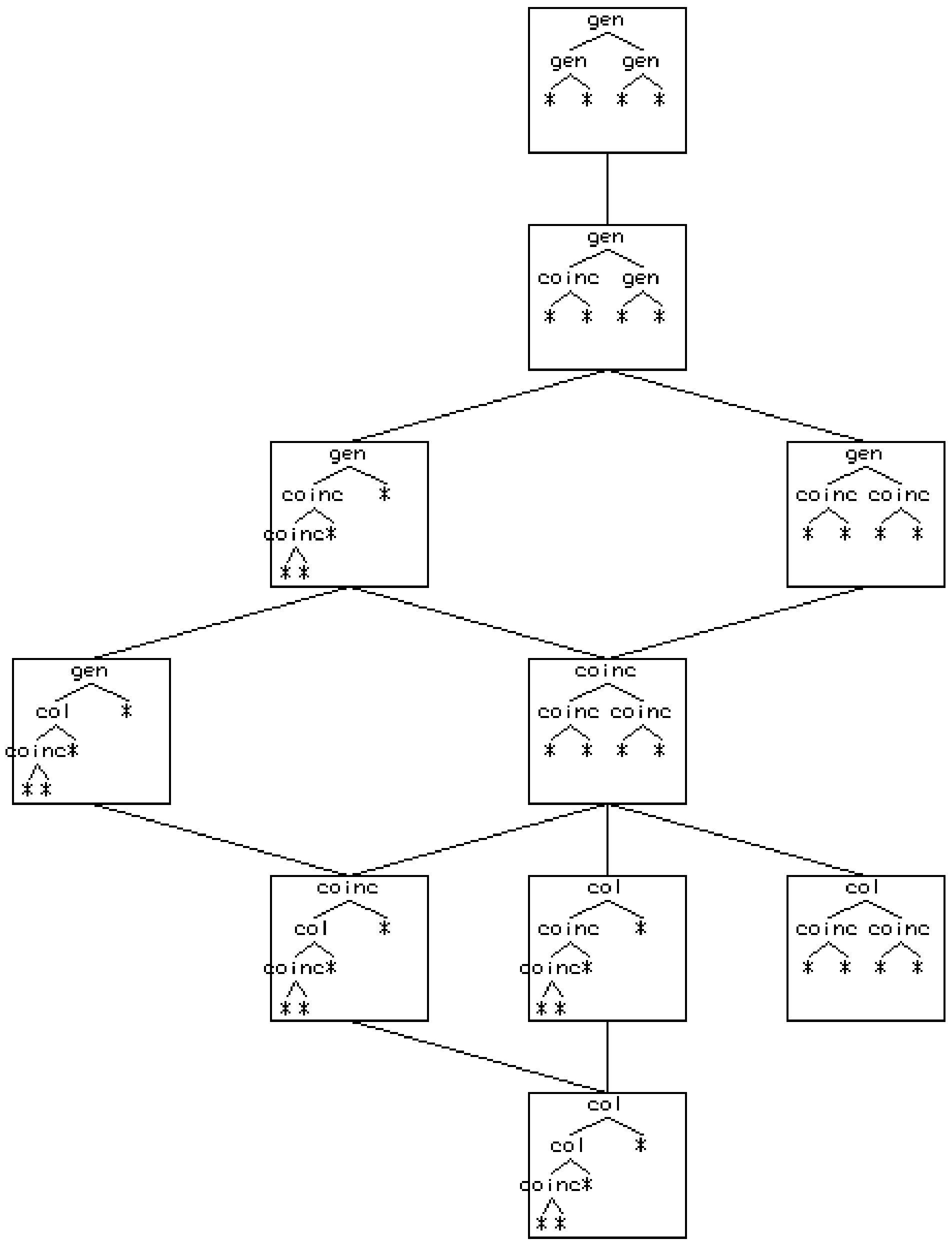


Figure 17. The “collapsed” space for $n = 4$.

Ben-Av & Sagi, 1995). *Closure* of curves is another qualitative regularity that human observers seem to respect (Kovacs & Julesz, 1993).

More generally, because the theory is specified at the logical rather than the algorithmic level, the entire Dots-World regularity set can easily be rewritten or replaced entirely, without altering the logic of inference. Experiments have been conducted using a regularity set suitable for use with edge fragments instead of dots, so that the grouping machinery can operate directly on the output of an edge finder. The interpretation of line drawings (including occlusion, 3-D pose, etc.) poses a far greater challenge than does simple grouping, but again can be approached using the same logical machinery augmented only by a suitable regularity set. This could potentially lead to a system for producing qualitative interpretations of line drawings with the same desirable properties of robustness, intuitiveness, and semantic optimality as are exhibited by the qualitative parse. This possibility is the subject of ongoing research.

Efficient implementation

The machinery described above is pitched strictly at the theoretical level, with no thought given to practicality. For example every possible partition of a given number of dots, of which there are an exponential number, was considered. This is in the very nature of an attempt to define the "best" interpretation, which is necessarily defined over the entire set of possible interpretations. The goal of the theory, that is, was to *define* the best interpretation, not to indicate the fastest way to find it.

Having laid the theory out in pure form, though, it is natural to turn to pragmatics. It turns out to be possible to design an efficient (worst case $O(n^2)$) approximation to the ideal solution. This approximation, which was used to construct the examples given above, is extremely faithful to the ideal solution. The implementation is described only briefly here; a more detailed investigation of the properties and performance of the algorithm is deferred to a separate paper.

The algorithm

Procedural vs. logical interpretation of lattices. Above, the partial orders $\overset{M}{\leq}$ and $\overset{T}{\leq}$ have been regarded *logically*, that is, as preference orders among logical forms. Alternatively, the same constructions can be regarded *procedurally*, as the sequence in which computations are to be carried out. This interpretation, combined with control over backtracking in Logic Programming using cuts, leads immediately to an efficient procedure for finding maximum-codimension interpretations.

The procedure is recursive. To find the interpretation of the dot configuration $x_0 \cup \mathbf{x}$,

1. find the best interpretation T of $\mathbf{x}$, and then
2. join x_0 to some subtree of T.

The recursion bottoms out in the minimal case of two dots, which are assigned an interpretation by means of clauses ordered in reverse order of codimension:

$\texttt{interpret}(\{x_1,x_2\})$ as $[\texttt{coincident},[x_1,x_2]] \leftarrow$ $\texttt{coincident}(x_1,x_2),!.$

[otherwise:]

$\texttt{interpret}(\{x_1,x_2\})$ as $[\texttt{generic},[x_1,x_2]]$.

Notice that these two clauses are a procedural instantiation of the $n=2$ lattice (Fig. 12).

Similarly, each new dot x_0 is added to the interpretation of $\mathbf{x}$ by means of ordered clauses. For each new dot x_0, the program generates subtrees of the existing interpretation, and then attempts to find a regular relationship between x_0 and each subtree, testing for higher-codimension relationships first (e.g., with the Dots World regularity set, starting with `collinear`). The procedure terminates as soon as x_0 fits generically in an existing subtree, and then pops up to the next level.

The order of the clauses ensures that both the lattice partial order and the genericity constraint are respected, because higher-codimension interpretations are always evaluated before lower-codimension interpretations.[9] This means that by the time an interpretation T succeeds, all more regular interpretations—which must *not* hold in order for T to be generic—have *already* failed.

Complexity

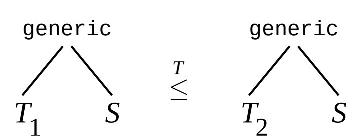
Because the procedure terminates with success but continues on failure (which corresponds to lack of structure in the configuration) processing time is highly dependent on the degree of structure in the configuration. The worst case is n mutually generic dots. In this case the program attempts to join the $n+1$-th dot with each of the existing n dots, for a total of $1+2+\ldots+n = O(n^2)$ processing steps. At the other extreme, a completely ordered configuration such as a collinear chain will take only n steps, as each new dot is added in one step to the single existing subtree. Hence the procedure is efficient in general, and extremely rapid when the image structure is salient. In practice the procedure takes only a few seconds on a Sparc Station 2 even on randomly generated displays of hundreds of dots.

Local approximations to the qualitative parse

Efficient as the above implementation is, it still computes an inherently *global* interpretation, because the preference order is defined over entire interpretations. Nevertheless, it is quite evident that grouping is highly dependent on local structure.[10] Hence it is worth noting that there is good reason to hope that the qualitative parse might be recoverable by local methods akin to those in the conventional grouping literature, perhaps providing an even more efficient computation.

This hope is provided by the internal organization of the space of parses, and in particular by its distributivity—coupled with the fact that in the Dots World context, a distant dot must be generic with respect to extant structure, because `coincident` is a precondition of `collinear`. The same would be true in any context in each regularity implies `coincident` under ω. As a result, mutually distant regions of a configuration can be regarded independently, in that $T_1 \overset{T}{\leq} T_2$ implies

(45)



regardless of S. A locally better solution is, a fortiori, a globally better one. Similarly, consider a configuration $\mathbf{x}$, which has qualitative parse $T_{\mathbf{x}}$. If a new dot x_0 is introduced that is mutually generic with each dot in $\mathbf{x}$, then it follows

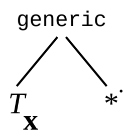
---

[9] The order of interpretations having the same codimension is unimportant, because the there is always a unique model in which the new dot fits generically.

[10] Here of course by the term "local" we mean local in the plane, rather than local in the partial order as was meant above.

immediately that the new configuration $\mathbf{x} \cup x_0$ has qualitative parse

(46)



Note these simple locality relationships might not hold for other contexts, such as one appropriate for line drawings. This accords with what one would expect psychologically, as such contexts are well known to have sometimes striking non-local effects. One might imagine that multiple percepts of impoverished scene configurations might correspond to multiple maximum-codimension solutions, drawn from a context definition corresponding to human regularity concepts (again see Richards et al., 1996).

Relevance to human vision

Despite its perhaps imposing computational formalism, the logical theory bears a close relationship to human perceptual organization. It is widely recognized that human higher-level vision often exhibits a certain "qualitative" character, wherein each potential interpretation is contraposed against distinctly different alternative interpretations. This is especially obvious with ambiguous figures, such as the Rubin vase/face figure, the Necker cube, or, to take an example from dot grouping, a square grid of dots that is perceived as either a set of horizontal strips or a set of vertical strips. In each of these figures there seem to be two distinct, equally plausible "solutions." But the point holds in the more typical situation where there is a single unique winning interpretation, chosen among qualitatively distinct alternatives. The observer must (unconsciously) choose among a possibly large set of alternative models of the scene—each of which, in one way or the other, accounts for the observed image. The process of deciding among these alternatives has often been described as one of quasi-logical inference and problem-solving (Rock, 1983), although the "reasoning" tools available are obviously more restricted than in more general cognitive contexts.

A particularly important—and particularly mysterious—component of the perceptual system is the process by which local qualitative decisions are combined with one another to form a globally preferred interpretation. In computational vision some progress was made in the 1970s on line-labeling in polygonal scenes, in which local qualitative decisions about lines and vertices propagate through the image (see Mackworth, 1976 for a review). This work foundered not only because the process proved inefficient but also because it was not accompanied by a well-motivated inference theory. In the psychological literature, it is known that local grouping inferences (e.g. the detection of local orientation) are aggregated into complexes that are much larger than the receptive field of any one simple cell (Dodwell, 1983; Field, Hayes, & Hess, 1993), a phenomenon sometimes referred to as "cooperativity" (Kubovy & Wagemans, 1995). Indeed, human observers have been shown to depend heavily on non-local structure in organizing the scene (Wagemans, Gool, Swinnen, & Horebeek, 1993). A key idea appears to be lateral agreement between distinct cues suggesting a common distal explanation (Enns & Rensink, 1991)—an idea closely related to codimension.

Thus the problem addressed by the lattice theory—how individual perceptual "atoms" are composed to form global interpretations—is a puzzle of long standing in psychology. Moreover the way preference among these interpretations is computed—via a regularity-based partial order—resonates with traditional approaches to the problem. The idea that the visual system chooses the most regular solution is ubiquitous, though controversial (Kanizsa, 1979), though the great subtlety inherent in the idea of "regularity" has generally defeated attempts to define it formally. The notion has its roots in the Gestalt notion of Prägnanz (good form) or the "minimum principle"—choose the simplest interpretation possible (see Hatfield & Epstein, 1985 for a review of minimum principles). Indeed, the partial order exhibited above can be regarded as an explicit mathematically realization of this idea.

Probably the most sophisticated modern incarnation of the minimum principle comes from the work of Leeuwenberg and colleagues (Leeuwenberg, 1971; Buffart, Leeuwenberg, & Restle, 1981; Van der Helm & Leeuwenberg, 1991), which builds on earlier attempts to define the "complexity" of a line drawing (e.g. see Hochberg & McAlister, 1953). In this work, figures are first described in a fixed description language, using the shortest description possible. Moreover, when choosing between two figures (e.g., when evaluating potential completions of a partially occluded figure), that with the shorter description is preferred. The results seem to agree well with human intuitions, for example correctly predicting when human observers will complete a partly occluded figure vs. when they will draw the "mosaic" interpretation (no occlusion).

This minimum rule is obviously related to the maximum-codimension rule proposed above, and the two approaches to some extent share a common spirit. Yet the lattice theory has a number of advantages. First, it is fully computable, whereas the coded descriptions described in the above papers were generated manually by the scientists. Second, it is well-motivated; the maximum-codimension rule derives from the Genericity Constraint, which in turn derives from the straightforward Bayesian goal of minimizing false inferences. By contrast, Leeuwenberg and colleagues themselves admit some mystery about why the minimum rule in Coding Theory actually works (though they have many informal insights). Finally, and most importantly, while Coding Theory provides a preference ordering on interpretations, it does *not* provide the interpretations; these must be provided by the researcher, which means by unanalyzed human intuitions. For example, while the "mosaic" interpretation mentioned above can be coded directly, the occlusion interpretation is, by definition, unavailable. In their study the authors simply code what they consider to be the intuitive completion, and compare its code to that of the mosaic interpretation in order to predict which is preferred. Even if the Coding Theory's minimum rule is correct, then, a major element of a rigorous account is missing. By contrast the lattice theory not only compares but *enumerates* interpretations; indeed the preference rule and the interpretations on which it operates emerge in tandem from the same formalism, the lattice.

Admittedly, the clean, simple world of dot grouping is vary far from the real world of noisy natural images. Indeed, it might seem difficult to imagine how a clean formalism like the lattice theory might apply in a more realistic setting. The weakest and most over-simplified link in the theory as presented above lies in the very impoverished definitions of the regularity predicates. The human visual system's definitions of collinearity and proximity, for example, exhibit a subtlety and context-sensitivity not captured by the simple thresholds used above. More realistic definitions are required, perhaps supplemented with a richer notion of "satisfaction" then the simple binary one used above. For example, in natural scenes the observer often must decide whether a particular noisy or textured region of the image is well-described as a single surface. Some progress has been made (e.g. in regularization theory) in defining the degree of fit between the data and a given surface model. However, conventional models lack a sufficiently rich notion of preference among models, such as that provided by suitable lattice. Combining the lattice preference order with a more sophisticated notion of degree of fit might provide some of the flexibility and power exhibited by human observers.

Conclusion: a competence theory for grouping

The regularity-based interpretation theory is presented as a competence theory (Clowes, 1971; Marr, 1982; Richards, 1988): an account of *what* interpretations are preferred and *why*, rather than of *how* they are computed—although the efficient implementation presented above demonstrates that the "how" follows rather easily. The highlight of the theory is the exhaustive enumeration of perceptual interpretations, which is demonstrably complete with respect to a given Closed World. Inference hinges on the partial order over these interpretations, in which preferred interpretations are always minima. Given a fixed set of distinguished regularity concepts, there is a unique interpretation in which an observed configuration is generic. In the hierarchical extension derived above, preference devolves to the maximum codimension grouping interpretation, the qualitative parse.

As Reiter and Mackworth (1989) argued in their original paper, the logical approach has the advantage that "interpretations" are given a full-fledged logical definition, suitable for proving semantic optimality and other desirable inferential properties. The current approach augments this rigorous definition of interpretations with a psychologically-motivated preference ordering—an idea Reiter and Mackworth explicitly invited. One might even imagine a well-defined idealized solution such as that proposed here might ultimately replace seat-of-the-pants perceptual inspection by human observers as the touchstone for evaluating the qualitative success of novel grouping methods.

Perhaps the most intriguing element here is the demonstration that classical high-level perceptual problems such as grouping are vulnerable to attack by powerful and well-understood techniques of Logic Programming. As suggested above, the approach seems particularly apt in domains characterized by *qualitative* perceptual interpretations of scene configurations. Such situations are are well known to be ubiquitous, but tend to be resistant to traditional approaches. Qualitative effects have often been viewed as stemming from non-accidental properties and other types of special configurations; the research reported above shows these perceptual primitives can serve as the atoms of a logical language with a well-defined inference structure. Moreover, this approach may permit a bridge between perceptual theory and the entire universe of model-theoretic computational semantics (Baral & Gelfond, 1994), a connection that is crucial to the future development of a "perceptual semantics."

References

Apt, K. (1990). Logic programming. In J. van Leeuwen (Ed.), *Handbook of theoretical computer science.* Elsevier.

Baral, C., & Gelfond, M. (1994). Logic programming and knowledge representation. *J. Logic Programming*, *19*(20), 73–148.

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

Bennett, B. M., Hoffman, D. D., & Prakash, C. (1989). *Observer mechanics: A formal theory of perception.* London: Academic Press.

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

Brookes, A., & Stevens, K. A. (1991). Symbolic grouping versus simple cell models. *Biological Cybernetics*, *65*, 375–380.

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

Caelli, T. M., & Julesz, B. (1978). On perceptual analyzers underlying visual texture discrimination: Part i. *Biological Cybernetics*, *28*, 167–175.

Clark, K. (1978). Negation as failure. In H. Gallaire & J. Minker (Eds.), *Logic and data bases.* New York: Plenum Press.

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

Cox, I. J., Rehg, J. M., & Hingorani, S. (1993). A bayesian multiple hypothesis approach to contour segmentation. *International Journal of Computer Vision*, *11*, 5–24.

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

Dodwell, P. C. (1983). The lie transformation group model of visual perception. *Perception and Psychophysics*, *34*, 1–16.

Enns, J. T., & Rensink, R. A. (1991). Preattentive recovery of three-dimensional orientation from line drawings. *Psychological Review*, *98*(3), 335–351.

Erné, M. (1993). Distributive laws for concept lattices. *Algebra Universalis*, *30*, 538–580.

Feldman, J. (1991). Perceptual simplicity and modes of structural generation. In *Proceedings of the 13th annual conference of the cognitive science society.*

Feldman, J. (1992a). Constructing perceptual categories. In *Proceedings of the ieee conference on computer vision and pattern recognition.*

Feldman, J. (1992b). *Perceptual categories and world regularities.* Unpublished doctoral dissertation, Massachussetts Institute of Technology. (Available as Technical Report #6, Rutgers Center for Cognitive Science)

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). The structure of perceptual categories. *Journal of Mathematical Psychology*, *41*, 145–170.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local ”association field”. *Vision Research*, *33*(2), 173–193.

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

Grätzer, G. (1978). *General lattice theory.* Basel: Birkhäuser.

Guy, G., & Medioni, G. (1992). Perceptual grouping using global saliency-enhancing operators. In *11th international conference on pattern recognition, vol. i.* The Hague, Netherlands.

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

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

Jacobs, D. W. (1989). *Grouping for recognition* (A.I. Memo No. 1177). Massachusetts Institute of Technology Artificial Intelligence Laboratory.

Jepson, A., & Richards, W. A. (1991). *What is a percept?* (Occasional Paper No. 43). MIT Center for Cognitive Science.

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

Jepson, A., & Richards, W. A. (1992b). A lattice framework for integrating vision modules. *IEEE Transactions on Systems, Man, and Cybernetics*, *22*(5), 1087–1096.

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

Kovacs, I., & Julesz, B. (1993). A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation. *Proceedings of the National Academy of Science*, *90*, 7495–7497.

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

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

Leyton, M. (1984). Perceptual organization as nested control. *Biological Cybernetics*, *51*, 141–153.

Leyton, M. (1988). A process-grammar for shape. *Artificial Intelligence*, *34*, 213–247.

Leyton, M. (1992). *Symmetry, causality, mind.* Cambridge, MA: M.I.T. Press.

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.

Mackworth, A. (1976). Model-driven interpretation in intelligent vision systems. *Perception*, *5*, 349–370.

Mackworth, A. K. (1988). Adequacy criteria for visual knowledge representation. In Z. Pylyshyn (Ed.), *Computational processes in human vision: an interdisciplinary perspective* (pp. 462–474). Norwood, N.J.: Ablex.

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

Parent, P., & Zucker, S. W. (1989). Trace inference, curvature consistency, and curve detection. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *11*(8), 823–839.

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

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

Reiter, R. (1978). On closed world data bases. In H. Gallaire & J. Minker (Eds.), *Logic and data bases.* New York: Plenum 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: M.I.T. Press.

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

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

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

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.

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

Witkin, A. P., & Tenenbaum, J. M. (1983). On the role of structure in vision. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and machine vision* (pp. 481–543). Academic Press.

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 similarity in dot patterns. *Perception and Psychophysics*, *34*(6), 513–522.