

Formation of visual “objects” in the early computation of spatial relations

JACOB FELDMAN

Rutgers University, New Brunswick, New Jersey

Perceptual grouping is the process by which elements in the visual image are aggregated into larger and more complex structures, i.e., “objects.” This paper reports a study of the spatial factors and time-course of the development of objects over the course of the first few hundred milliseconds of visual processing. The methodology uses the now well-established idea of an “object benefit” for certain kinds of tasks (here, faster within-object than between-objects probe comparisons) to test what the visual system in fact treats as an object at each point during processing. The study tested line segment pairs in a wide variety of spatial configurations at a range of exposure times, in each case measuring the strength of perceptual grouping as reflected in the magnitude of the object benefit. Factors tested included nonaccidental properties such as collinearity, cotermination, and parallelism; contour relatability; Gestalt factors such as symmetry and skew symmetry, and several others, all tested at fine (25 msec) time-slices over the course of processing. The data provide detailed information about the comparative strength of these factors in inducing grouping at each point in processing. The result is a vivid picture of the chronology of object formation, as objects progressively coalesce, with fully bound visual objects completed by about 200 msec of processing.

The organization of the initially inchoate visual field into coherent units or “objects,” called perceptual grouping or binding, is of fundamental importance in visual perception, influencing the perception of lightness (Adelson, 1993; Gilchrist, 1977), motion (Shimojo, Silverman, & Nakayama, 1988), and recognition of objects (Biederman, 1987). Much has been learned about the grouping factors originally identified by the Gestaltists, such as spatial proximity (Compton & Logan, 1993; Kubovy & Wagemans, 1995), collinearity (Caelli & Umansky, 1976; Claessens, & Wagemans, 2005; Feldman, 1997a, 2001; Pizlo, Salach-Golyska, & Rosenfeld, 1997; Smits, Vos, & Oeffelen, 1985), and parametric similarity (Geisler & Super, 2000; Zucker, Stevens, & Sander, 1983).

Yet surprisingly little is known about the role of more complex geometric factors on grouping, notwithstanding a large number of conflicting theoretical proposals. One particularly important and often-discussed case is grouping of line segments, because simple oriented elements are thought to be the basic units in the early visual system. But even the minimal case of two line segments is poorly understood. What spatial relations between two line segments tend to induce binding? Many answers to this question have been proposed, with surprisingly little direct psychophysical evidence to adjudicate. One very influential proposal is that grouped configurations tend to be *nonaccidental properties* (Lowe, 1987; Witkin & Tenenbaum, 1983), that is, configurations that seem unlikely to have occurred by accident—or, in Horace Bar-

low’s phrase, “suspicious coincidences” (Barlow, 1994). Such configurations, including (at least) *collinearity* (Figure 1A), *cotermination* (Figure 1B), and *parallelism* (Figure 1C), probabilistically imply the presence of stable structure in the 3-D world (Jepson & Richards, 1992), and hence make promising candidates for grouping. Biederman (1987) provided some initial empirical evidence that nonaccidental groupings are perceptually important and highly detectable. However more recent experimental investigations (Wagemans, 1992) have yielded more equivocal results, leaving open the question of whether nonaccidental properties actually do in fact promote grouping, and if so, to what degree; and how the binding induced by the three best-known nonaccidental properties compare to each other.

Moreover, several other geometric factors have been proposed as candidate grouping cues. Prominent among these are realizations of the Gestalt notion of good continuation, including *cocircularity* (Parent & Zucker, 1989; Ullman, 1976) and *relatability* (Kellman & Shipley, 1991) (Figure 1D). (These two relations approximately coincide in the case of line segment pairs, and in what follows I will use the term relatability to encompass either one.) Two line segments are relatable if a smooth curve can be passed from the end of one to the beginning of the other without loops or inflections (Singh & Hoffman, 1999). Relatability is not strictly equivalent to any nonaccidental property, though it is of course closely related to collinearity. It is thought to play a major role in early visual completion (Field, Hayes,

J. Feldman, jacob@rucss.rutgers.edu

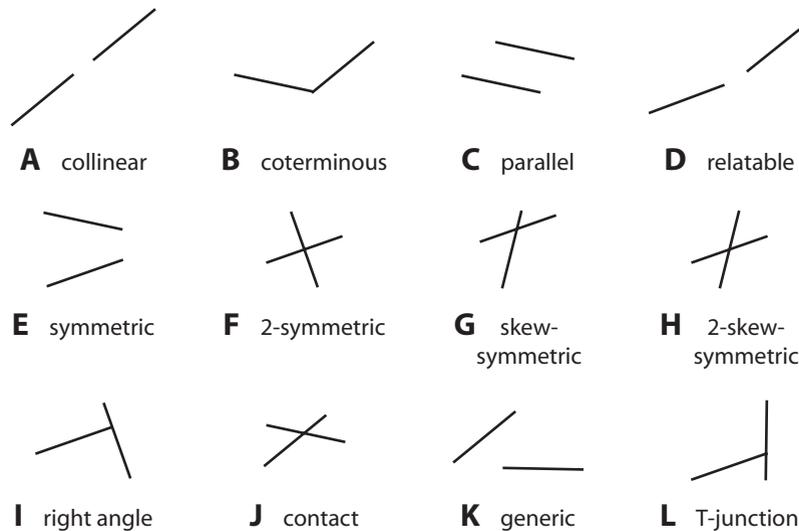


Figure 1. Illustrations of some of the spatial relations discussed in the text.

& Hess, 1993), but its tendency to promote grouping of line segments, and in particular relative to other potential grouping cues, is not well known.

The field of potential grouping factors expands considerably more when one considers more global or configural factors such as bilateral symmetry (Figure 1E), biaxial mirror symmetry (Figure 1F) skew symmetry (i.e., bilateral symmetry as viewed from a slant; Figure 1G), and even biaxial skew symmetry (Figure 1H). Symmetry (Wagemans, 1995, 1993) and other kinds of global regularity (Boselie & Wouterlood, 1989; Kanizsa, 1979; Leeuwenberg, 1971) have long been invoked as organizational principles in the visual system, and have even been found to override local cues when the two are put in conflict (Sekuler, Palmer, & Flynn, 1994). However local cues are more amenable to early computation by the visual system, raising the question of how their relative time-courses of local and global cues might differ.

Several other types of line segment configurations are of particular interest. *Perpendicularity* (Figure 1I) is another special configuration that has been thought to have an early role in assigning local 3-D structure (Enns & Rensink, 1991). One also naturally wonders whether actual contact (Figure 1J) matters, as would be predicted by the principle of “uniform connectedness” (Palmer & Rock, 1994) (but not corroborated by recent findings [Kimchi, 2000])—but not if nonaccidentalness were the sole factor, because transverse crossing of lines (as in the figure) is not nonaccidental.

Contraposed to all of these regular and potentially binding-inducing configurations is the case of totally irregular or *generic* line segment pairs (Figure 1K) (“none of the above” cases), which bear no apparent structural relationship. Such configurations constitute the natural baseline against which more regular candidate configurations ought to be compared.

Finally, one particularly interesting question is the case of *T-junctions* (Figure 1L), which are thought to signal

occlusion boundaries (Clowes, 1971). These are in effect an *anti*-grouping cue, in that the head and stem of the T would normally be interpreted as, respectively the occluding and occluded boundaries of two distinct objects (though of course in particular situations other interpretations are possible). Hence the question of whether T-junctions induce binding is particularly urgent.

To complicate matters further, many of the above-mentioned categories overlap, raising the question of how multiple cues might combine in influencing binding, a difficult question recently of great theoretical interest (Feldman, 1997b; Landy, Maloney, Johnston, & Young, 1995). Hence a thorough investigation would include not only cases of each of the above properties but would range over the entire space of possible line segment pairs, including all geometrically possible combinations of the relations of interest.

Chronology of Perceptual Grouping

The time-course of perceptual grouping is also very little understood. Debate exists over whether grouping is preattentive, and thus presumably early (Prinzmetal & Banks, 1977; Treisman, 1982), or postattentive and thus presumably late (Mack, Tang, Tuma, Kahn, & Rock, 1992). Simple scene classification may be substantially complete by 150 msec of processing (Thorpe, Fize, & Marlot, 1996), though some visual illusions do not become completely effective until as much as 250 msec of exposure (Reynolds, 1978). Subjective contours have been found to be computed as early as about 100 msec (Ringach & Shapley, 1996). Amodal completion (i.e., completion of a contour behind a visible occluder) is computed early (Rensink & Enns, 1998), being substantially complete by 200 msec of processing (Sekuler & Palmer, 1992). A more recent study with more precise methodology indicates processing as early as 75–100 msec (Murray, Sekuler, & Bennett, 2001). An important contribution are the studies of Kimchi (1998, 2000), which have charted the time-course

of early grouping processes using objective methodology (primed matching), although not directly testing the contribution of the various two-segment geometric factors considered here. These diverse findings raise the possibility that grouping might involve a number of different rules or mechanisms, each with their own timing profiles, making it difficult to map out the chronology without a temporally fine-grained methodology and systematic manipulation of cues.

The current paper reports the results of a large experiment testing the degree of binding induced by a wide variety of line segment configurations, including examples of all of the above-mentioned special categories and combinations. The experiment used a new psychophysical methodology capable of very accurate and chronologically precise estimates of the magnitude of binding. Measurements of grouping strength for each configuration type were taken at a finely sampled sequence of time-slices, yielding an unprecedentedly detailed look at the chronology of perceptual grouping in perhaps its most basic case.

The Experimental Paradigm

Historically, experiments on visual grouping have often used “subjective” methods, in which subjects are asked for conscious judgments of how stimuli are grouped. Such methods limit the temporal precision with which binding estimates can be made, because they reflect only the subject’s final judgment, and hence potentially tap cognitive and decision procedures subsequent to visual processing as well as early visual processes themselves.

To obtain a more “objective” and precise estimate of the time-course of grouping processes, this study uses a method suggested by recent findings in the literature on object-based attention (Baylis, 1994; Duncan, 1984). Comparisons of visual features are more rapid and accurate *within* a perceptual object than *between* distinct objects (Behrmann, Zemel, & Mozer, 1998), a finding sometimes called the *object benefit*.¹ The idea is to turn this effect around and use it to determine what exactly is perceived as a visual object—that is, what is perceptually grouped or bound. To do this, we define a target configuration—whose structure will be manipulated—and ask the subject to execute a perceptual comparison between visual elements located either (1) within the target configuration or (2) between the target configuration and another, distinct object (called the foil). As we manipulate the structure of the target configuration, we change what grouping cues it obeys, and thus influence the degree to which it is, in fact, bound together as a single “object.” The resulting object effect (*within-target-configuration* RT minus *between-objects*RT) then acts as a measure of the degree to which the target configuration was grouped or bound together, and thus reflects the effectiveness of the grouping cues in effect in the particular target configuration. When using this differential measure, more negative differentials (within responses faster than between) indicate stronger grouping of the target configuration.

The idea that the within-object response-time advantage in this task can be taken as a measure of grouping is supported by findings that object-based attentional effects

extend to objects that are the result of perceptual grouping and completion processes. As mentioned, Behrmann et al. (1998) found that the object benefit transfers to objects completed behind an occluder, and several papers (Moore, Yantis, & Vaughan, 1998; Pratt & Sekuler, 2001) have shown that a similar object benefit (in a cued-comparison task) applies to objects created by both amodal completion and modal completion. These studies suggest that the “objects” in object-based attention are products of the same organizational processes that have traditionally been studied in the context of perceptual grouping (see Scholl, 2001 for discussion); certainly this is the universal assumption in the object-based attention literature.² Hence the purpose of the current study is not to corroborate that visual comparisons are faster within grouped objects—that is taken as a premise—but rather, obversely, to use this tendency as a tool for measuring the extent to which particular configurations are, in fact, perceptually grouped.

In the experiment described below, the subject’s task was to compare two probe dots (located either *within* the target configuration or *between* it and the foil) with respect to size (same or different); see Figure 2. Latency to evaluate such a comparison is known to depend in various ways on the intervening structure between the two probe locations. Studies of visual tracing along a curve have found the speed of this process to be influenced by the shape of the intervening curve (Jolicœur, Ullman, & Mackay, 1986, 1991). Roelfsema, Scholte, and Spekrijse (1999) found increased latency when probe comparisons crossed an intervening curve. Barenholtz and Feldman (2003) found that latency to make probe comparisons depends on both the magnitude and sign of intervening curvature; comparisons are slower along more sharply curved contours, and in particular along negatively curved (concave) ones—apparently because these tend to create perceptual part boundaries. This last finding in particular suggests that probe comparisons can be used to assess the degree to which the two probe locations fall in what are perceived to be perceptually distinct organizational units in the image.

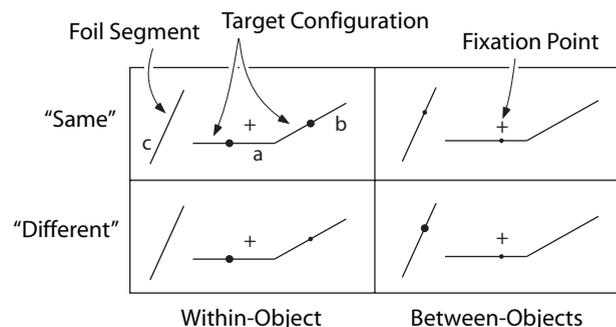


Figure 2. Examples of stimulus configurations, showing within-target-configuration (i.e., within potential object, left), between-objects (right), “same” response (top) and “different” response (bottom). The fixation cross is shown before the trial begins but is not visible during the trial.

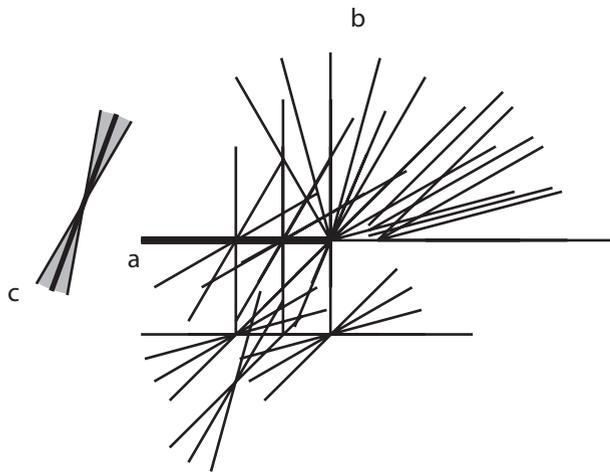


Figure 3. Diagram of configurations tested. Segment a is the always drawn at a standard position (heavy bar). Segment b varied as shown, determining the structure of the configuration. Segment c was the standard foil object, always drawn within the positions shown in gray. a-b comparisons are within-target-configuration a-c are between the target configuration and the foil.

METHOD

In the experiment, two line segments (each about $3.8^\circ \times .04^\circ$ of visual angle in size) serve as the target configuration (labeled a and b in Figure 2). The spatial arrangement of this pair was chosen from a wide range of possible configurations (Figure 3) including examples of many candidate grouping cues. In addition, a third segment c was present, located at a “generic” location assumed to induce only a baseline level of grouping. This configuration of three segments was presented on the computer screen for a variable interval (stimulus onset asynchrony, SOA), and then two small probe dots appeared superimposed on them. The dots appeared either on segments a and

b (*within* case) or on segments a and c (*between* case). After a brief interval (200 msec., brief enough to prevent the initiation of most eye movements), the stimulus and probes were replaced by a mask consisting of many random dots of both sizes.

The subject's task was to indicate whether the two probe dots were the same or different in size. Dots were either small (diameter $.15^\circ$) or large ($.23^\circ$). The two dots were either both the same size (same trials) or one small and one large, determined randomly (different trials). The screen distance between dots was the same on all trials (about 2.3° of visual angle) regardless of spatial configuration or any other factors, thus ensuring that variations in reaction time were due only to the influence of the geometry of the underlying line configuration. The entire three-line configuration was randomly rotated on each trial, always centered at the fixation point (indicated by a cross prior to the beginning of the trial, but not visible during the trial; see Figure 2). Segment a was always near the center of the three-line configuration, so centering the configuration at fixation guaranteed that segments b and c were always approximately equally eccentric. The 45 configurations were crossed with within/between objects (2 levels) and same/different probe sizes (2 levels) and 6 levels of SOA for a total of 1,080 trials per subject, in randomized order. Lines and probes appeared white on a black screen at high contrast in a darkened room. The subject's head position was fixed by a chinrest at 45-cm viewing distance.

Subjects

Subjects were run in two series, identical except for the levels of SOA used. An initial group of 58 subjects were run on a relatively coarse set (0, 200, 400, 600, 800, and 1,000 msec). Inspection of their data suggested that grouping asymptoted by about 200 msec (see results below), so a second group of 68 subjects was run on a finer set of SOAs (50, 75, 100, 125, 150, and 175 msec). Data from both groups are combined in the plots below because the number of subjects was large enough to create only negligible differences mean differences between the two groups. Subjects performing below 75% correct on the same/different task were removed from the analysis, leaving 43 and 58 subjects for analysis in the two groups, respectively. Trials with reaction times more than two standard deviations above each subject's mean were also removed.

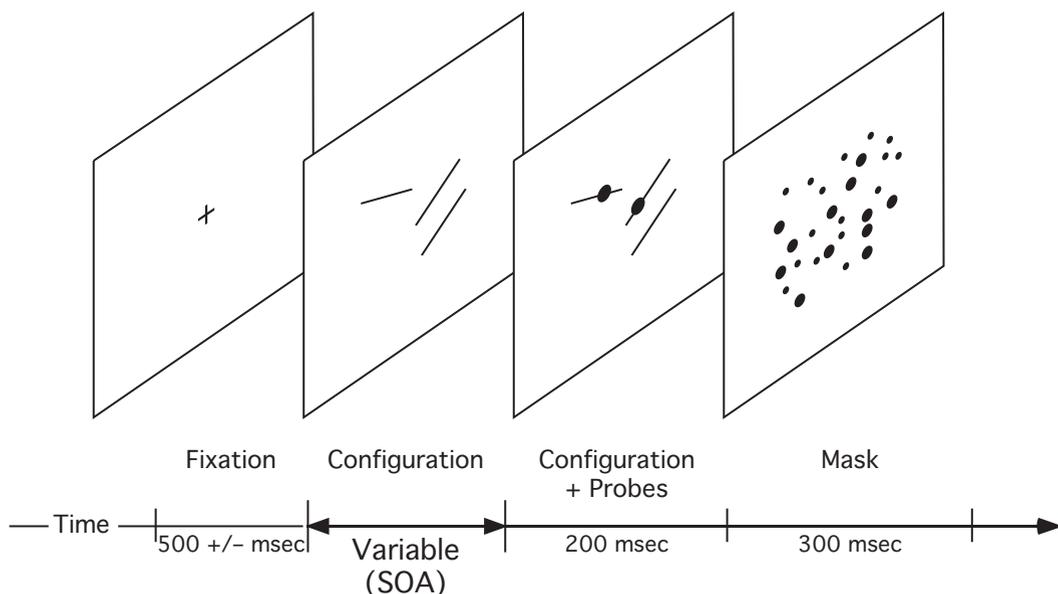


Figure 4. Sequence of frames on each trial: fixation, the configuration, including segments a, b, and c (see Figure 2); the same configuration with probe dots added; and a mask.

Configurations Tested

A set of 45 distinct configurations was constructed (Figure 3), intended to span virtually the complete space of possible geometric relations between two line segments, while including both examples and nonexamples of every relation of particular interest discussed above, including collinear, coterminous, parallel, symmetric, doubly symmetric, skew symmetric, doubly skew symmetric, T-junction, perpendicular, contacting, as well as generic cases (i.e., cases exhibiting “none of the above”). Figure 3 shows the full set schematically (with the position of segment a held fixed, while b varies to illustrate the relation). Of course the many features of potential interest are not independently crossed in the experiment, because (as discussed above) many of them are not geometrically independent (e.g., coterminous segments are always in contact). The foil segment c was always near a particular fixed generic location (shown in the figure), but in order to avoid “overlearning” of its relative placement, its exact orientation was jittered randomly (within 10°, range shown in gray) about the indicated mean orientation.

Although the spatial distance between the probe locations was always fixed, the distance between the centers of the segments on which they lay varied somewhat from configuration to configuration as required by the geometrical relations desired. Hence the proximity of the segments can also be regarded as a potential grouping cue, and will be evaluated as such in the analyses below, but it should be emphasized that the variation in proximity was a side-effect of these other geometric factors, and furthermore was necessarily constrained so that the probe locations could lie at the required distance; so the resulting manipulation of distance was small in magnitude and should not be regarded as a serious test of the effect of proximity on perceptual grouping.

RESULTS AND DISCUSSION

As discussed, the main dependent measure in the analysis is the within-between response time differential, i.e., the within-object benefit. Negative differentials indicate an object benefit, with lower (more negative) numbers indicating stronger binding. The primary analysis strategy is simply to extract a configuration subset of interest (e.g., all parallel configurations), and plot its differential RT as a function of SOA. Such a plot shows the strength of the object effect as a function of the presumed duration of processing.

Figure 5 shows plots of the time-varying object effect for nonaccidental (panel A), relatable (B), symmetric and skew symmetric (C), perpendicular (D), contacting (E), and T-junction (F) cases. Most of these plots clearly show a progressively increasing object effect (diminishing differentials), with each plot beginning at a high level (usually about +60 msec), and gradually dropping to an asymptotic low level after about 200 msec of processing. (At long exposures, especially after 200 msec when eye movements are possible, the data begin to exhibit more variance, and seem generally to regress to 0; hence the ensuing discussion will mostly be restricted to the first 200 msec of processing.) These plots vividly show the progressive development of “objects”—that is, configurations bound strongly enough to exhibit an object benefit—as the visual system executes its analysis of the spatial relations present in the image. Grouping computations begin (at 0 msec of processing) with an unbound object, and eventually terminate with a maximally bound object by about 200 msec of processing.

Figure 5G shows a summary plot of the effects of the eight factors (excluding T-junctions, discussed below) whose raw data is shown the panels A–F. This plot shows for each factor the best fit exponential decay function, allowing easy comparisons of the binding effect due to the various factors. (As discussed below, the T-junction data are excluded from this summary because they do not fit the exponential model well.) The factors generally have similar time-constants (all about $\beta = -0.015$, or a half-life of about 50 msec), but differ only in their baselines, indicating that processing of all cues takes approximately the same amount of time, but with some cues imparting stronger ultimate binding effects than others, or, equivalently, imparting more binding effect per millisecond of processing than others. The summary plot clusters the curves into approximately significantly distinct bins by their asymptotes (the bins are shown within gray bands in the plot).

Comparing Grouping Factors

In considering the relative grouping strength of the various factors, though, one must consider that as discussed above they are heavily correlated with one another, as they would be too in natural images. In order to factor out the true influence of each factor on grouping, all the cues were entered into a multivariate linear regression with object effect (RT differential) as the dependent variable. Here in order to derive the most information from the predictors, cues were rendered in numerical rather than binary form wherever possible: e.g., the number of symmetry axes (0, 1, or 2) or skew symmetry axes (0, 1, or 2). In addition, in order to profit from the variance accounted for by simple SOA, a linearization of SOA ($\exp[-0.015 \times \text{SOA}]$) was entered in the regression.

The results of the regression, shown in Table 1, indicate that the following factors significantly induce binding, listed in order of their statistical influence: relatability, the number of mirror symmetry axes, and parallelism. After partial covariations are compensated for, the number of skew symmetric axes, perpendicularity, and cotermination have *negative* influence on binding. At first blush this latter finding is surprising, but reflects the nature of the correlations among cues. For example coterminating config-

Table 1
Results of the Regression Testing the Influence and Significance of Various Grouping Factors on the Object Benefit

| Factor | Relative Standard Regression Weight | <i>p</i> |
|-------------------|-------------------------------------|----------|
| Lin. SOA | .2349 | <.0001 |
| Relatable | -.1036 | .0019 |
| # sym. axes | -.0909 | .0053 |
| # skew sym. axes | .1702 | <.0001 |
| Coterminating | .0947 | .0031 |
| Parallel | -.0863 | .0001 |
| Right angle | .0993 | <.0001 |
| Collinear | -.0389 | .1638 |
| Centroid distance | -.0304 | .3124 |
| Contact | -.0508 | .1391 |

Note—Negative weights indicate positive binding effects.

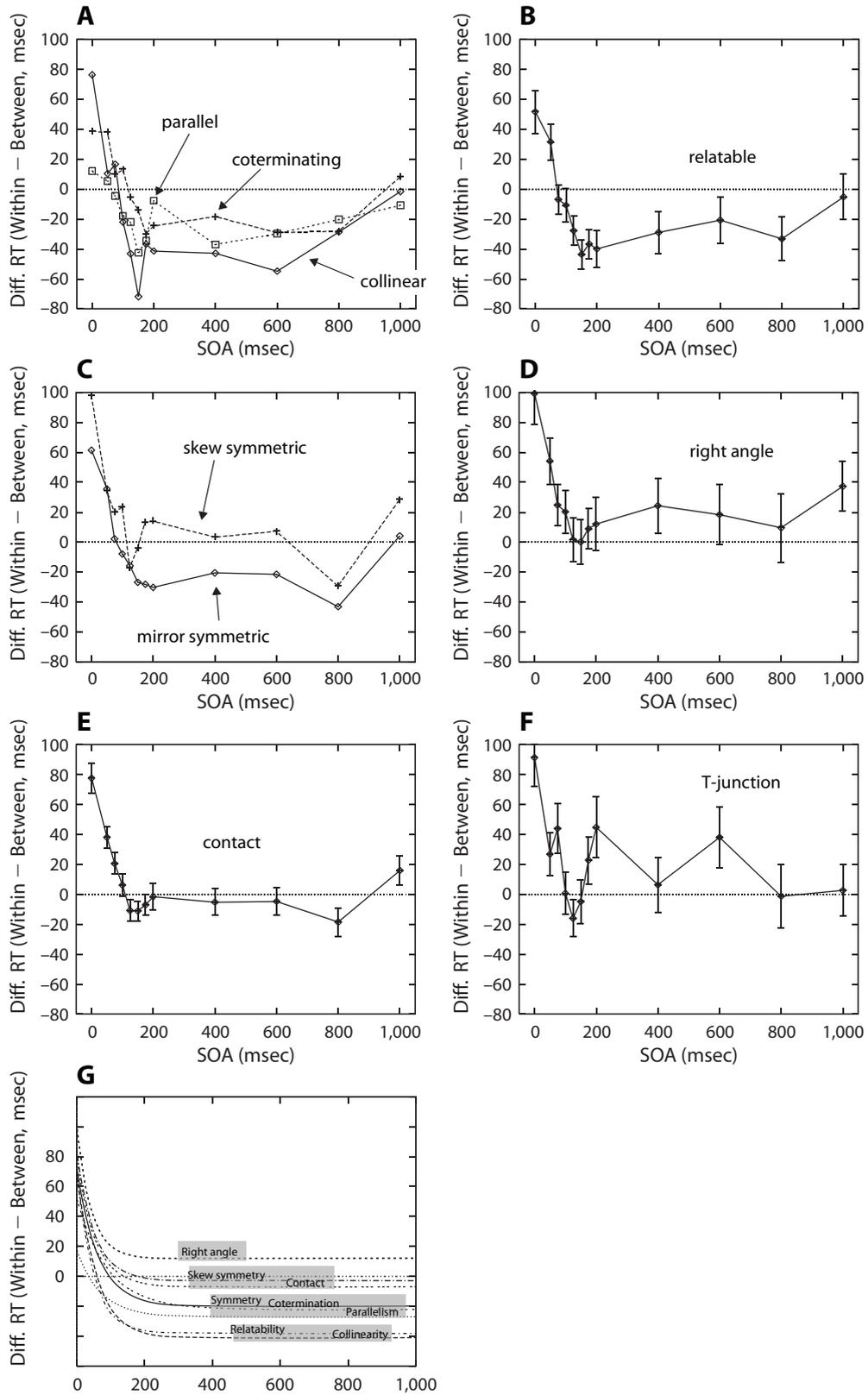


Figure 5. Results of the experiment, showing the object effect (ordinate) for selected subsets of the configurations as a function of processing time (SOA, abscissa). Most of the plots clearly show a progressive development of the object effect over time, asymptoting by about 200 msec of processing. Panel G shows for each configuration type the best fit exponential decay function (model: $\text{Differential RT} = \text{baseline} + a \times \exp[\beta \times \text{SOA}]$), allowing easier comparisons of the time course and strength of the various grouping factors.

urations are always relatable, but when the angle is large make poor binding; while noncoterminating but relatable configurations with shallow angles bind well, leading to a paradoxical negative influence of cotermination after its partial correlation with relatability has been factored out. A similar account explains the negative effect of skew symmetry, which correlates heavily with ordinary mirror symmetry but is at best a weaker binding cue. Finally, contact, proximity, and collinearity had no significant effect on grouping.³ Again collinear cases (which strictly require 0° deviation in the two segments' orientations) *do* have a strong probinding effect (as is plain in Figure 5A), but it is almost completely accounted for by the substantially correlated factor of relatability.

A second analysis investigated interactions among the grouping factors. As mentioned, the factors tested are not all orthogonal, neither in principle (due to geometry) nor in the specific set of configurations tested. Hence not all combinations of factors can be simultaneously considered. A quasi-independent (not orthogonal) subset of the factors, chosen to be as large as possible, was entered in a second regression along with all their pairwise interactions (products). Of these, two exerted significant influence on grouping (RT differential): relatability \times the number of symmetry axes ($p < .002$); and the number of skew symmetry axes \times right angle ($p < .05$). In both cases, the regression coefficient was positive, meaning that the combined factor *negatively* impacted grouping. (Recall that a grouping effect is signaled by a negative regression coefficient, meaning that the factor made the more RT differential more negative.) Thus in each of these two pairs of grouping factors, the two factors were in effect "laterally inhibiting" each other, each having less influence in the presence of the other than they would otherwise. Again though this finding should be interpreted cautiously, only implying that within the narrow geometric constraints of the chosen configurations, grouping strength as measured in the experiment tended to saturate, so that when one was strong the other no longer had a clearly measurable effect. Certainly in a more naturalistic setting one would imagine grouping factors to interact in a mutually "excitatory" manner.

In summary, these data suggest that grouping is indeed induced by regular spatial relations among line segments: in particular: (1) "good-continuation" cues, (2) symmetry, and (3) parallelism, with several of these factors interacting subadditively. Simple contact—whether two segments touched or not—had no effect, contrary to the principle of uniform connectedness (Palmer & Rock, 1994), but consistent with the ubiquitous experience of visual items that overlap in the image but do not perceptually cohere due to orientation structure, like the branches of a tree and the roof of the house behind them. Proximity of the target line segments also showed no measurable grouping in these data, but as mentioned above no strong conclusions should be drawn from this, as the segments' proximity was only varied over a very small range compared to the much more substantial variation in the other grouping factors mentioned.

One conspicuous feature of these plots requires comment. The initial (SOA = 0) differential is not generally

0 msec (zero object benefit), as one might expect: instead it usually begins at some positive value, generally near +60 msec, before beginning to drop. This means that between trials, which involve the fixed foil, are actually faster than within trials until a substantial degree of configuration-specific binding has taken place. This in turn presumably reflects the extreme familiarity of the foil configuration, which is repeated on every trial, i.e., a kind of "overlearning"—an account corroborated by the recent finding that familiarity and learning can exert a strong, early influence on grouping (Kimchi & Hadad, 2002; Zemel, Behrmann, Mozer, & Bavelier, 2002). This learning-dependent level can be taken as a kind of pseudozero, representing the average within-between differential prior to the influence of grouping factors. Because the foil is fixed, the absolute level of grouping with it (i.e., between segments a and c) is unimportant, as long as grouping within the target configuration is compared with this fixed baseline. For this reason the binding effects reported in this paper should all be taken as reflecting *relative* binding strength among grouping cues (and among SOAs), rather than absolute degrees of binding.

Relative Orientation

The relative orientation of the two segments plays an evident role in the definitions of some of the factors discussed above, such as collinearity, parallelism, and perpendicularity, in each of which classes relative orientation is fixed. More subtly, relative orientation may also make a more graded contribution as it varies within a configuration class. A case in point is relatability, which in its original formulation by Kellman and Shipley (1991) is strictly dichotomous (relatable at less than 90°, nonrelatable otherwise). Hence it is particularly interesting to look at the effect of angle within relatable cases.

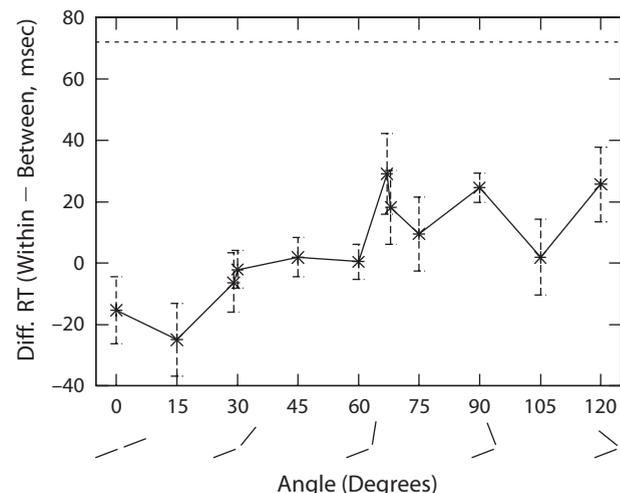


Figure 6. Object benefit (within – between RT) as a function of absolute angular deviation between the two line segments for relatable cases, showing a steady decrease in grouping effect with increasingly obtuse angles. The dotted line indicates the "pseudozero" (see text for discussion).

Figure 6 shows the differential (within-between) object effect as a function of relative orientation (absolute value) for reliable cases only. As might be expected, the effect is maximal (largest negative differential) at near collinear (0°) cases, and diminishes steadily at larger angles (linear regression [$F(1,23882) = 26.5568, p < .000001$]. The gradual reduction in binding at progressively more obtuse angles, rather than abrupt fall-off at some angular threshold, is consistent with recent work by Guttman, Sekuler, and Kellman (2003) (and in fact also by Kellman and Shipley (1991) themselves, who found a gradual reduction in the proportion of subjects reporting subjective contours as angle increased; see their Figure 28). This agreement provides some incidental validation of the novel methodology employed here.

The Influence of Regularity on Grouping

The above findings suggest that a number of geometric features contribute to grouping, all consistent with the mental construction of an organized whole. How do these factors combine computationally? Many authors have suggested that perceptual grouping realizes some kind of process of complexity-minimization or regularity-maximization (e.g., Hochberg & McAlister, 1953; Leeuwenberg, 1971; see Hatfield & Epstein, 1985 for a review). One computational realization of this idea is *minimal model* theory (Feldman, 1997b, 2003a), which attempts to define and select the least “accidental” interpretation of a given image, i.e., the one with the fewest unexplained coincidences (cf. Rock, 1983). To a good approximation, this maximum-codimension interpretation corresponds to what human observers see in a large configuration of dots or oriented edges, overtly pulling out salient groups, curves, and “objects.”

Briefly, the idea behind minimal model theory is to choose the grouping interpretation that, given a configuration, maximizes regularity, given quantitatively by the *codimension*. This term, which derives from standard use in analytic geometry, refers to the number of degrees of freedom that are fixed in a configuration relative to the generic case (Poston & Stewart, 1978). Its relevance to perception was first proposed by Jepson and Richards (1992) who pointed out that it can be used to represent the degree of structural regularity exhibited by a given visual configuration, in that it encodes the number of special configurational properties (“regularities”) that the configuration obeys relative to a totally generic or free configuration.

Codimension thus can serve as a numeric representation of the degree of regularity characteristic of a particular configuration such as a line segment pair. By definition, generic line pairs (Figure 1K) will have codimension 0, parallel codimension 1 (because one parameter, angle, is fixed relative to the generic case), coterminating codimension 2 (because the two-dimensional location of one endpoint is forced), etc. The codimension can be thought of as the degree of “nesting” in the hierarchy of regularity definitions: e.g., parallel is a special case of general position (relative angle set to 0), collinear is a special case of parallel (offset set to 0), and so forth. A vivid way

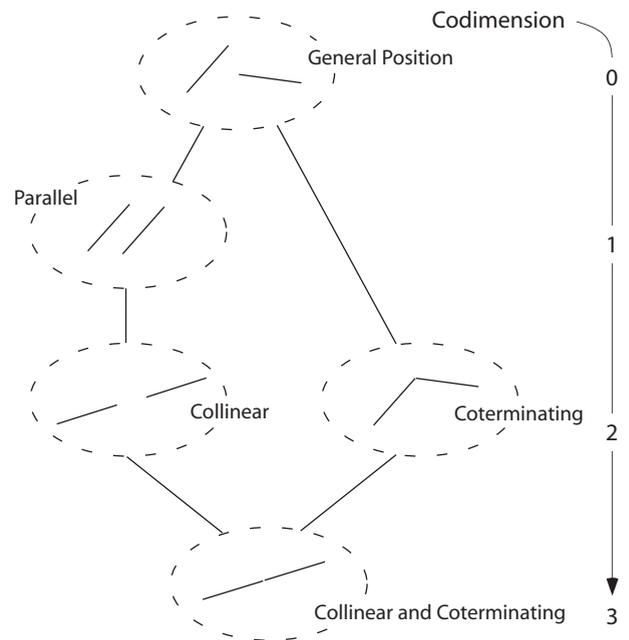


Figure 7. Lattice of configurations as partially ordered by regularity (see Feldman, 1997). Codimension (right column) corresponds to the row number, counted down from zero at the top.

of understanding this is to literally diagram these nesting relationships, which form a lattice indicating the inclusion relations among the predicates (Feldman, 1997b, 2003a). In this kind of diagram codimension is simply the “row number” of the configuration class in question, starting with zero at the top, indicating how many levels of regularity configurations in that class obey. Figure 7 shows such a lattice for the two-segment configurations used in the current study, with codimensions indicated.

A convenient aspect of minimal model theory is that perceptual “objects” correspond to *subtrees* of the parse tree that describes the perceptual organization of the field of visual elements. Moreover, the “strength” or salience of an object corresponds to the codimension of the corresponding subtree, which measures its internal regularity and thus how strongly separated it is from the rest of the tree (Feldman, 1999, 2003b). This observation entails a very direct prediction for the current experiment: that the object benefit accruing to each configuration will increase monotonically with its codimension. The contrast between within-object and between-objects comparison paths, as viewed through the inferred tree description, is illustrated schematically in Figure 8. Probe comparisons within the target configuration (a-b) must traverse only the narrow subtree corresponding to the target configuration; while between-objects probe comparisons (a-c) must traverse a longer path spanning the entire tree. The target configuration’s codimension in effect modulates how strongly this tree interpretation is assigned to the scene, and thus how strongly this asymmetry between the two comparison paths is realized.

Figure 9 plots codimension against within-between differential, collapsing over all other factors and SOA. The

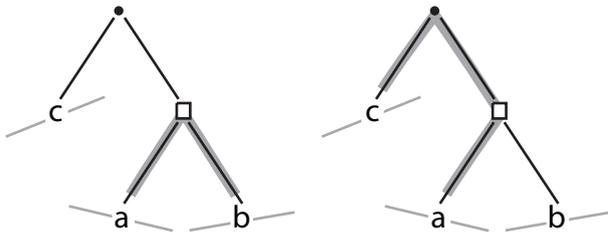


Figure 8. Schematic tree diagrams (parse trees) illustrating the difference between the paths (heavy gray curves) that the comparison must traverse in a-b (within-target-configuration, left) versus a-c (between-objects, right) probe comparisons. (See Feldman, 1997b, 2003a, for fuller explanation of trees like these.) A within-target-configuration comparison traverses a small subtree (corresponding to the induced “object”), whereas a between-objects comparison must traverse a relatively long path comprising an additional node. The □ symbol indicates the node manipulated in the experiments; the nature of the spatial relation at this node modulates how strongly the configuration will be assigned this tree.

plot clearly shows the predicted increase in object benefit with codimension [linear regression $F(1,41862) = 14.4691, p = .0001$]. As codimension (regularity) increases, configurations are bound more tightly into objects. The measurement of codimension also relates directly to probabilistic or Bayesian models of perceptual organization (Feldman, 2001; Geisler, Perry, Super, & Gallogly, 2001; Kersten, Mamassian, & Yuille, 2004). From a Bayesian point of view, nonaccidental and other “special” spatial relations can be identified with hypotheses that have elevated prior probability (Jepson & Richards, 1992). The lattice-based choice rule from minimal model theory is analytically equivalent to Bayes’ rule if priors are suitably assigned—specifically, if nodes on the lattice (e.g., the one in Figure 7) are assigned equal priors (Feldman, under review). Loosely, each nonaccidental (or other special) property can be thought of as having

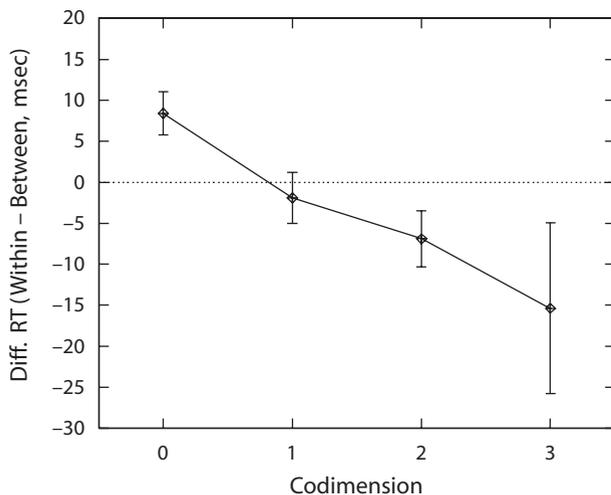


Figure 9. As codimension (regularity) increases, objects are bound more strongly.

some small probability ϵ of occurring “by accident,” but occurs with probability 1 if the property is a stable aspect of the scene. This leads to a likelihood ratio (and under the assumption of equal priors, a posterior ratio) of $1/\epsilon$ in favor of the more regular interpretation. The codimension is the *number* of such properties in effect, so for a configuration with codimension C , the likelihood (and posterior) ratio climbs rapidly, as $1/\epsilon^C$. This in turn means that the Bayesian evidence (log likelihood ratio) in favor of the “object” interpretation increases linearly with codimension C (because $\log[1/\epsilon^C] = \log[\epsilon^{-C}] = C \log [1/\epsilon]$). This can be thought of as the Bayesian interpretation of the linear increase in binding effect (increasingly negative RT differential) plainly visible in Figure 9. But whether viewed through Bayes or through minimal model theory, the interpretation is conceptually the same: both say that the strength of the regular interpretation of the configuration—its degree of “objecthood”—increases with its degree of internal regularity, whether measured by the likelihood ratio in its favor (Bayes) or, equivalently, by its depth in the lattice (minimal model theory).

The Case of T-Junctions

As discussed above, the case of T-junctions is particularly interesting, because such configurations suggest occlusion boundaries and thus two *distinct* objects. As can be seen in Figure 5F, T-junctions constitute an exception to the general trend of a monotonically increasing object effect. Here the usual pattern of decreasing differentials obtains for the first 125 msec of processing, but then reverses, climbing back to about +40 msec by 200 msec of SOA when grouping has generally completed. This rise, monotonic through 125, 150, 175, and 200 msec of processing, is significant by linear regression [$F(1,2154) = 8.4182, p = .0038$]. After 200 msec, the differential begins to regress back toward 0, as with most other cases. Notice that while a small rise in same interval is visible in several of the other plots (e.g., right angle, contact, and skew symmetry), these overlap substantially with T-junctions; thus the rise is being driven by the T-junctions, where it is much more salient.

The interpretation of this rise is straightforward: after an initial period (0–125 msec) of grouping, T-junctions undergo a subsequent period of *ungrouping* (125–200 msec), as the initially bound object begins to disaggregate back into its two component line segments. This pattern is especially interesting for the light it sheds on the temporal dynamics of local versus global cue competition in grouping. Presumably the initial grouping is due to local provisional cues, such as the centroid proximity or contact of the two line segments, while the later revised interpretation of distinct objects is based on a more holistic analysis of the entire configuration. Local cues dominate at first, but configurational cues prevail in the end, with the competition resolved by 200 msec. A similar early local versus global competition has been found in the case of visual motion estimation (Pack, Berezovskii, & Born, 2001). This finding also relates closely to that of Rauschenberger and Yantis (2001), who similarly found

that an early grouping interpretation—not consistent with what eventually reaches awareness—is briefly in play during the first 200 msec of computation before apparently being suppressed.

The Progressive Development of Objects: A Chronological Picture

The above discussion analyzed the data with respect to given, predetermined features of interest. The progress of object formation can be appreciated in a different and more theory-neutral way by simply drawing all individual configurations that have achieved a significant object benefit at a given SOA (Figure 10). Each such picture, referred to as a “hotmap,” shows exactly what configurations the visual system has bound by each given point in the processing stream. The development of these hotmaps over time constitutes a kind of time-lapse photographic montage of the progressive formation of objects.

The most immediate conclusion from the hotmap sequence is that the number of bound configurations generally increases with SOA, consistent with the generally monotonically increasing strength of binding revealed by the earlier plots. At SOA = 0, the only bound objects are simple collinear and relatable configurations, consistent with the sensitivity of early visual cortex to oriented image structure. At SOA = 50, more relatable cases come in, as well as some parallel and symmetric cases. By later SOAs, more and more cases with a wider variety of structures appear. By SOA = 200, relatable, symmetric, and

parallel cases dominate, especially in the most significant cases (shown in black). Notably, at SOA = 125, several T-junctions have bound, but by SOA = 200 all of them have disbanded and thus disappeared.

CONCLUSIONS

The data reported here paint a picture of perceptual object formation in which grouping proceeds rapidly and progressively over the first 200 msec of processing. During this time the object benefit exhibited by grouped configurations increases measurably and monotonically. Binding is ultimately strongest in configurations exhibiting regular spatial relations of certain types, principally including good continuation (relatability or collinearity), parallelism, and symmetry. No single feature seems to be essential to grouping, which rather depends on the general degree of regularity and structure exhibited by the configuration, as captured numerically by its codimension.

The data also provide answers to several previously unanswered questions about the rules governing grouping. Nonaccidental properties do induce grouping, but only with some caveats: collinearity induces grouping only in the more general guise of relatability and Gestalt good continuation, and cotermination induces grouping only when the angle is small, again consistent with good continuation. Symmetry, by its nature a less local and more configurational cue, induces grouping even after short processing times, reflecting its importance in the

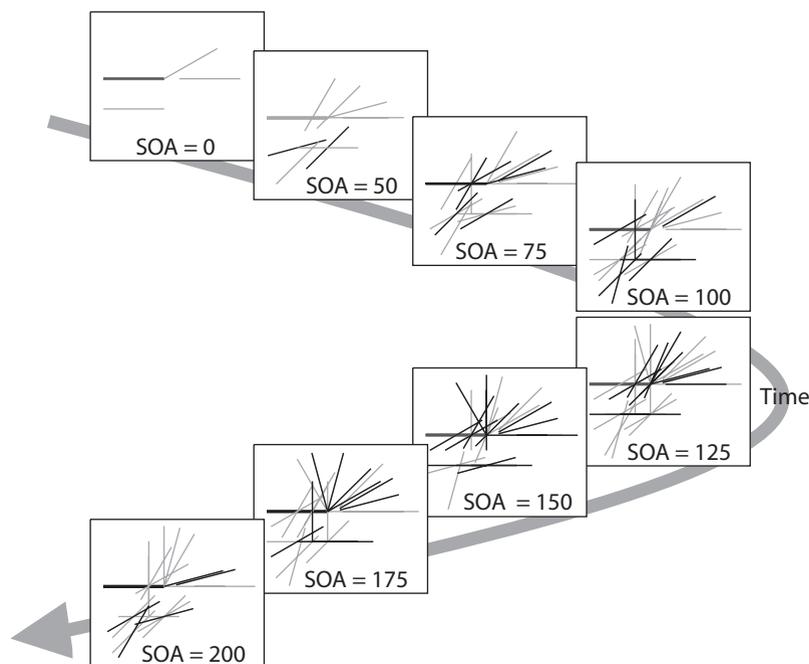


Figure 10. “Hotmaps,” i.e., illustrations of the configurations exhibiting an object effect at each time slice from SOA = 0 to 200 msec. Each picture shows the target configurations (drawn in the same standard coordinate frame as in Figure 3) whose within-between differential was significantly faster than the pseudozero ($p < .01$ shown in gray, $p < .0001$ in black). These pictures thus literally show the progressive development of visual “objects” over time.

early computation of pattern structure (Kukkonen, Foster, Wood, Wagemans, & van Gool, 1996; Wagemans, 1997; Wagemans, van Gool, Lamote, & Foster, 2000). The interplay of cues across time is revealed particularly vividly by the data for T-junctions, which suggest that local cues dominate early while configurational cues eventually override them. A more complete computational theory of grouping will be required before we can understand more fully the mechanisms underlying this competition, especially with more complex configurations than those tested here.

AUTHOR NOTE

I am grateful to Cordelia Aitkin, Elan Barenholtz, David Fass, Whitman Richards, and Manish Singh for helpful discussions, and to Ed Kuang for assistance in data collection. Support was provided by NSF Grant SBR-9875175 and NIH (NEI) Grant EY15888. Please direct correspondence to J. Feldman, Department of Psychology, Center for Cognitive Science, Rutgers University, 152 Frelinghuysen Road, Piscataway, NJ 08854 (e-mail: jacob@ruccs.rutgers.edu).

REFERENCES

- ADELSON, E. H. (1993). Perceptual organization and the judgment of brightness. *Science*, **262**, 2042-2044.
- BARENHOLTZ, E., & FELDMAN, J. (2003). Visual comparisons within and between object parts: Evidence for a single-part superiority effect. *Vision Research*, **43**, 1655-1666.
- BARLOW, H. (1994). What is the computational goal of the neocortex? In C. Koch & J. L. Davis (Eds.), *Large-scale neuronal theories of the brain* (pp. 1-22). Cambridge: MIT Press.
- BAYLIS, G. C. (1994). Visual attention and objects: Two-object cost with equal convexity. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 208-212.
- BEHRMANN, M., ZEMEL, R. S., & MOZER, M. C. (1998). Object-based attention and occlusion: Evidence from normal participants and a computational model. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 1011-1036.
- BIEDERMAN, I. (1987). Recognition by components: A theory of human image understanding. *Psychological Review*, **94**, 115-147.
- BOSELIE, F., & WOUTERLOOD, D. (1989). The minimum principle and visual pattern completion. *Psychological Research*, **51**, 93-101.
- CAELLI, T. M., & UMANSKY, J. (1976). Interpolation in the visual system. *Vision Research*, **16**, 1055-1060.
- CLAESSENS, P. M., & WAGEMANS, J. (2005). Perceptual grouping in Gabor lattices: Proximity and alignment. *Perception & Psychophysics*, **67**, 1446-1459.
- CLOWES, M. B. (1971). On seeing things. *Artificial Intelligence*, **2**, 79-116.
- COMPTON, B. J., & LOGAN, G. D. (1993). Evaluating a computational model of perceptual grouping by proximity. *Perception & Psychophysics*, **53**, 403-421.
- DAVIS, G., & HOLMES, A. (2005). Reversal of object-based benefits in visual attention. *Visual Cognition*, **12**, 817-846.
- DUNCAN, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, **113**, 501-517.
- ENNS, J. T., & RENSINK, R. A. (1991). Preattentive recovery of three-dimensional orientation from line drawings. *Psychological Review*, **98**, 335-351.
- FELDMAN, J. (1997a). Curvilinearity, covariance, and regularity in perceptual groups. *Vision Research*, **37**, 2835-2848.
- FELDMAN, J. (1997b). Regularity-based perceptual grouping. *Computational Intelligence*, **13**, 582-623.
- FELDMAN, J. (1999). The role of objects in perceptual grouping. *Acta Psychologica*, **102**, 137-163.
- FELDMAN, J. (2001). Bayesian contour integration. *Perception & Psychophysics*, **63**, 1171-1182.
- FELDMAN, J. (2003a). Perceptual grouping by selection of a logically minimal model. *International Journal of Computer Vision*, **55**, 5-25.
- FELDMAN, J. (2003b). What is a visual object? *Trends in Cognitive Sciences*, **7**, 252-256.
- FELDMAN, J. (2007). *Bayes and the simplicity principle in perception*. Manuscript under review.
- 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**, 173-193.
- GEISLER, W. S., PERRY, J. S., SUPER, B. J., & GALLOGLY, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, **41**, 711-724.
- GEISLER, W. S., & SUPER, B. J. (2000). Perceptual organization of two-dimensional patterns. *Psychological Review*, **107**, 677-708.
- GILCHRIST, A. L. (1977). Perceived lightness depends on perceived spatial arrangement. *Science*, **195**, 185-187.
- GUTTMAN, S. E., SEKULER, A. B., & KELLMAN, P. J. (2003). Temporal variations in visual completion: A reflection of spatial limits. *Journal of Experimental Psychology: Human Perception & Performance*, **29**, 1211-1227.
- HATFIELD, G., & EPSTEIN, W. (1985). The status of the minimum principle in the theoretical analysis of visual perception. *Psychological Bulletin*, **97**, 155-186.
- HOCHBERG, J., & MCALISTER, E. (1953). A quantitative approach to figural "goodness." *Journal of Experimental Psychology*, **46**, 361-364.
- JEPSON, A., & RICHARDS, W. A. (1992). What makes a good feature? In L. Harris & M. Jenkin (Eds.), *Spatial vision in humans and robots* (pp. 89-125). Cambridge: Cambridge University Press.
- JOLICŒUR, P., ULLMAN, S., & MACKAY, M. (1986). Curve tracing: A possible basic operation in the perception of spatial relations. *Memory & Cognition*, **14**, 129-140.
- JOLICŒUR, P., ULLMAN, S., & MACKAY, M. (1991). Visual curve tracing properties. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 997-1022.
- KANIZSA, G. (1979). *Organization in vision: Essays on Gestalt perception*. New York: Praeger.
- KELLMAN, P. J., & SHIPLEY, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, **23**, 141-221.
- KERSTEN, D., MAMASSIAN, P., & YUILLE, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, **55**, 271-304.
- KIMCHI, R. (1998). Uniform connectedness and grouping the perceptual organization of hierarchical patterns. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 1105-1118.
- KIMCHI, R. (2000). The perceptual organization of visual objects: A multigenetic analysis. *Vision Research*, **40**, 1333-1347.
- KIMCHI, R., & HADAD, B.-S. (2002). Influence of past experience on perceptual grouping. *Psychological Science*, **13**, 41-47.
- KUBOVY, M., & WAGEMANS, J. (1995). Grouping by proximity and multistability in dot lattices: A quantitative gestalt theory. *Psychological Science*, **6**, 225-234.
- KUKKONEN, H. T., FOSTER, D. H., WOOD, J. R., WAGEMANS, J., & VAN GOOL, L. (1996). Qualitative cues in the discrimination of affine-transformed minimal patterns. *Perception*, **25**, 195-206.
- LANDY, M. S., MALONEY, L. T., JOHNSTON, E. B., & YOUNG, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, **35**, 389-412.
- LEEUWENBERG, E. L. J. (1971). A perceptual coding language for visual and auditory patterns. *American Journal of Psychology*, **84**, 307-349.
- LOWE, D. G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, **31**, 355-395.
- MACK, A., TANG, B., TUMA, R., KAHN, S., & ROCK, I. (1992). Perceptual organization and attention. *Cognitive Psychology*, **24**, 475-501.
- MOORE, C. M., YANTIS, S., & VAUGHAN, B. (1998). Object-based visual selection: Evidence from perceptual completion. *Psychological Science*, **9**, 104-110.
- MURRAY, R. F., SEKULER, A. B., & BENNETT, P. J. (2001). Time course of amodal completion revealed by a shape discrimination task. *Psychonomic Bulletin & Review*, **8**, 713-720.
- PACK, C. C., BEREZOVSKII, V. K., & BORN, R. T. (2001). Dynamic properties of neurons in cortical area MT and anaesthetized macaque monkeys. *Nature*, **414**, 905-908.
- PALMER, S., & ROCK, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review*, **1**, 29-55.

- PARENT, P., & ZUCKER, S. W. (1989). Trace inference, curvature consistency, and curve detection. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **11**, 823-839.
- PIZLO, Z., SALACH-GOLYSKA, M., & ROSENFELD, A. (1997). Curve detection in a noisy image. *Vision Research*, **37**, 1217-1241.
- POSTON, T., & STEWART, I. N. (1978). *Catastrophe theory and its applications*. London: Pitman.
- PRATT, J., & SEKULER, A. B. (2001). The effects of occlusion and past experience on the allocation of object-based attention. *Psychonomic Bulletin & Review*, **8**, 721-727.
- PRINZMETAL, W., & BANKS, W. (1977). Good continuation affects visual detection. *Perception & Psychophysics*, **21**, 389-395.
- RAUSCHENBERGER, R., & YANTIS, S. (2001). Masking unveils pre-amodal completion representation in visual search. *Nature*, **410**, 369-372.
- RENSINK, R. A., & ENNS, J. T. (1998). Early completion of occluded objects. *Vision Research*, **38**, 2489-2505.
- REYNOLDS, R. I. (1978). The microgenetic development of the Ponzo and Zöllner illusions. *Perception & Psychophysics*, **23**, 6.
- RINGACH, D. L., & SHAPLEY, R. (1996). Spatial and temporal properties of illusory contours and amodal boundary completion. *Vision Research*, **36**, 3037-3050.
- ROCK, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press.
- ROELFSEMA, P. R., SCHOLTE, H. S., & SPEKREIJSE, H. (1999). Temporal constraints on the grouping of contour segments into spatially extended objects. *Journal of Experimental Psychology: Human Perception & Performance*, **39**, 1509-1529.
- SCHOLL, B. J. (2001). Objects and attention: The state of the art. *Cognition*, **80**, 1-46.
- SEKULER, A. B., & PALMER, S. E. (1992). Perception of partly occluded objects: a microgenetic analysis. *Journal of Experimental Psychology: Human Perception & Performance*, **121**, 95-111.
- SEKULER, A. B., PALMER, S. E., & FLYNN, C. (1994). Local and global processes in visual completion. *Psychological Science*, **5**, 260-267.
- SHIMOJO, S., SILVERMAN, G. H., & NAKAYAMA, K. (1988). An occlusion-related mechanism of depth perception based on motion and interocular sequence. *Nature*, **333**, 265-268.
- SINGH, M., & HOFFMAN, D. D. (1999). Completing visual contours: The relationship between reliability and minimizing inflections. *Perception & Psychophysics*, **61**, 943-951.
- SMITS, J. T., VOS, P. G., & OEFFELEN, M. P. VAN. (1985). The perception of a dotted line in noise: A model of good continuation and some experimental results. *Spatial Vision*, **1**, 163-177.
- THORPE, S., FIZE, D., & MARLOT, C. (1996). Speed of processing in the human visual system. *Nature*, **381**, 3.
- TREISMAN, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception & Performance*, **8**, 194-214.
- ULLMAN, S. (1976). Filling-in the gaps: The shape of subjective contours and a model for their generation. *Biological Cybernetics*, **25**, 1-6.
- WAGEMANS, J. (1992). Perceptual use of nonaccidental properties. *Canadian Journal of Psychology*, **46**, 236-279.
- WAGEMANS, J. (1993). Skewed symmetry: A nonaccidental property used to perceive visual forms. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 364-380.
- WAGEMANS, J. (1995). Detection of visual symmetries. *Spatial Vision*, **9**, 9-32.
- WAGEMANS, J. (1997). Characteristics and models of human symmetry detection. *Trends in Cognitive Sciences*, **1**, 346-352.
- WAGEMANS, J., VAN GOOL, L., LAMOTE, C., & FOSTER, D. H. (2000). Minimal information to determine affine shape equivalence. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 443-4468.
- 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). Orlando, FL: Academic Press.
- ZEMEL, R. S., BEHRMANN, M., MOZER, M. C., & BAVELIER, D. (2002). Experience-dependent perceptual grouping and object-based attention. *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 202-217.
- ZUCKER, S. W., STEVENS, K. A., & SANDER, P. (1983). The relation between proximity and brightness similarity in dot patterns. *Perception & Psychophysics*, **34**, 513-522.

NOTES

1. Davis and Holmes (2005) reported that when a variety of confounds are removed and stimulus variables suitably controlled, the within-object benefit actually reverses, with the advantage shifting to between-objects comparisons. The study reported below did not generally show such a reversal, and the conditions under which such a reversal would obtain are not perfectly clear. The issue is not critical in the current paper, as the logic of the methodology only requires that a within- versus between-object difference of *some kind* obtains, by which the formation of objects may be measured.
2. Indeed, to hypothesize otherwise—i.e., to envision two completely disjoint and parallel tracks of object creation mechanisms, one controlled by perceptual grouping and the other by attentional selection—seems singularly unparsimonious.
3. Of course, this should not be taken as contradicting the well-known effect of proximity on perceptual grouping more generally (see Kubovy & Wagemans, 1995). The noneffect here simply means that with spatially oriented objects such as line segments, proximity between the centroids of the two segments has no effect per se after the contributions of orientation-dependent geometric factors are removed.

(Manuscript received November 21, 2005;
revision accepted for publication December 18, 2006.)