

# Grouping by similarity is mediated by feature selection: evidence from the failure of cue combination

Liqiang Huang

Published online: 5 February 2015  
© Psychonomic Society, Inc. 2015

**Abstract** As one of the most classic gestalt phenomena, grouping is often considered to occur preattentively. Therefore, it seems natural to expect conflicting grouping cues to combine preattentively and to lead to an impression of “no grouping.” On the other hand, a recent account suggests that grouping by similarity is mediated by the attentional selection of a feature. This account makes the distinctive prediction that for grouping by similarity, but *not* for low-level groupings (e.g., connectedness, common region), when two conflicting grouping cues are present, the grouping structure will tend to be driven either by one type of cue or by the other type of cue and will usually not lead to the impression of “no grouping.” This predicted pattern of results was confirmed in the present study: the presence of two conflicting low-level grouping cues led to reports of “no grouping” in 73.0 % of the trials, but for two similarity grouping cues, “no grouping” was only reported in 25.5 % of the trials. In summary, although low-level grouping cues do indeed work together to determine the perceptual structure preattentively, grouping by similarity is likely mediated by the attentional selection of a feature.

**Keywords** Grouping · Grouping cues · Similarity · Attention · Feature selection

## Introduction

As one of the most classic gestalt phenomena, grouping is often considered to occur preattentively. However, two recent studies have suggested otherwise. Huang and Pashler (2007)

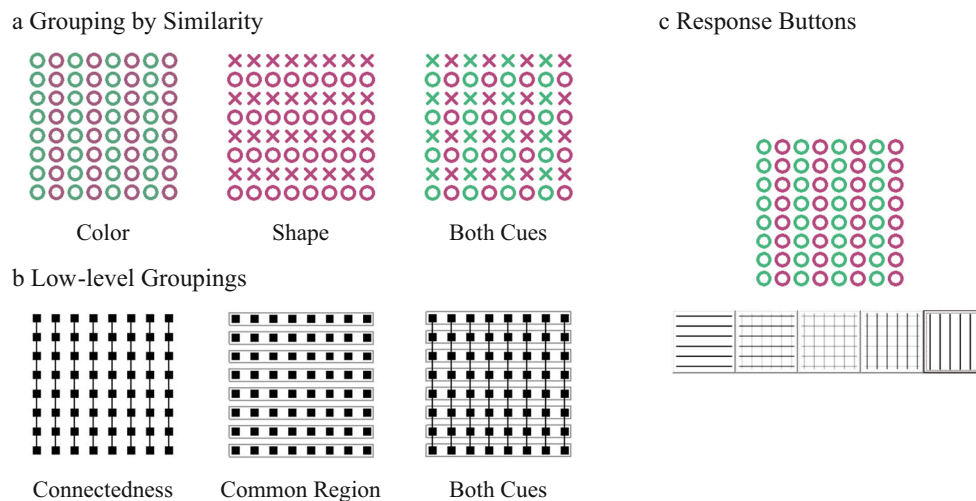
have argued that grouping by similarity is mediated by selective attention to individual features. When an observer selects an individual feature, the items will be organized through grouping by similarity, but when an observer tries to simultaneously select both features, the items will inevitably turn into a grouping by proximity. Thus, grouping by similarity seems to be mediated by feature selection in the sense that the grouping information is only available when a group is exclusively selected.

Levinthal and Franconeri (2011) provided critical support for this *feature selection account* of grouping. In their experiments, several pairs of dots moved around and the two dots of each pair moved together and formed a common-fate group. The search for a target pair was laborious, suggesting that information about such groups is not simultaneously available and has to be created by attending to the groups sequentially.

In this study, I further investigate the nature of grouping by similarity (being a preattentive process vs. being mediated by feature selection) in terms of the question of combinations of multiple grouping cues. A demonstration of this phenomenon is given in Fig. 1a: color cues will suggest grouping into columns, whereas shape cues will suggest grouping into rows. If, as suggested by Huang and Pashler (2007) and Levinthal and Franconeri (2011), grouping by similarity is mediated by feature selection, then two of these grouping cues could not combine into a “sum” because they are driven by attention to individual features. Instead, color and shape cues will compete for the control of attention and, assuming equal strength, will each dominate in half of the trials. Introspectively, the observers would feel that the perceptual structure of the “both cues” pattern in Fig. 1a is sometimes driven by a color-grouping cue and sometimes by a shape-grouping cue, and the percept would switch back and forth between these two possibilities. Phenomenally, this is similar to the bistable structures studied since Attneave (1971). Bistable structures demonstrate that these different perceptual interpretations do *not* combine

---

L. Huang (✉)  
Department of Psychology, The Chinese University of Hong Kong,  
Shatin, NT, Hong Kong SAR, The People’s Republic of China  
e-mail: lqhuang@psy.cuhk.edu.hk



**Fig. 1** The research question. **a** shows examples of the three types of stimuli displays for grouping by similarity: color-grouping cue, shape cue, and both cues. **b** shows examples of the three types of stimuli displays for low-level groupings: connectedness-grouping cue, common region-grouping cue, and both cues. Phenomenally, the “both-cues”

pattern in panel a tends to be driven either by the color-grouping cue or by the shape-grouping cue, and the percept would switch back and forth between these two possibilities. In contrast, the “both-cues” pattern in panel b tends to give an impression of “no grouping”. See text for details. **c** shows the response buttons

perceptually (i.e., preattentively), and the present study intends to make the same argument for grouping by similarity.

The feature selection account suggests that grouping by similarity and low-level groupings (e.g., grouping by connectedness, proximity, or common region; Palmer and Rock 1994; Rensink and Enns 1995; Palmer 1992) are fundamentally different from each other (Huang and Pashler 2007; Levinthal and Franconeri 2011). Unlike the former, low-level grouping cues are *not* mediated by feature selection, and it seems reasonable to assume a preattentive mechanism counts all the “votes” of the different low-level grouping cues. Therefore, two conflicting, and equally strong, low-level cues will frequently lead to a sum of approximate zero (i.e., no grouping on either direction). Thus, unlike grouping by similarity, I predict that when two low-level cues are both present in a pattern (e.g., connectedness and common region in the “both cues” pattern in Fig. 1b), this will indeed lead to the impression of “no grouping.” This “preattentive combination” mechanism will be elaborated below in the *General discussion* section.

In summary, the feature selection account predicts that a combination of two conflicting cues will lead to the impression of no grouping in low-level groupings but to a bistable structure driven by the two individual cues in grouping by similarity. Introspectively, the predictions of the feature selection account seem to be confirmed (see the “both cues” conditions in Fig. 1a and b).

## The present study

Unfortunately, the two possibilities illustrated above (bistable structure vs. no grouping) are indistinguishable in a

conventional task: a forced choice between two grouping structures. Both possibilities predict a split between the two grouping structures. In the present study, I added a new option to the responses: *absence of a clear grouping on either direction*. In the case of no grouping, the observers would frequently choose this new option. However, in the case of a bistable structure, the responses would be split between the individual grouping cues and the option of no grouping would not be chosen very frequently. For example, in Fig. 1a, the observers will attend to a color or a shape and neither will lead to the impression of “no grouping.”

Applying the above theoretical discussions in this specific design leads to the following predictions of the feature selection account of grouping by similarity: the “no-grouping” option will only be frequently used in low-level groupings, but not in grouping by similarity.

## Experiments

### Method

#### Participants

University undergraduate students, all of whom had normal or corrected-to-normal vision, participated in this study’s experiments. Three participants were excluded because they could not respond correctly to the conditions in which there was only one unambiguous cue (average accuracy on grouping direction <0.7). A further four participants were excluded, one because she almost never (<10 %) and three because they

almost always (>90 %) chose the “no grouping” option for “both-cues” conditions. Aside from these excluded participants, a total of 29 participants took part in the study. The inclusion of the excluded participants in the analysis would not change any of the conclusions of the present study.

### Apparatus

In both experiments, the stimuli were presented on a 1,024 × 768 pixels CRT monitor and the participants viewed the display from a distance of about 60 cm. The participants were asked to make responses by clicking one of five buttons (Fig. 1c). They were asked to respond as accurately as possible but were under no time pressure (i.e., unspeeded responses).

### Stimuli

Examples of the stimuli displays in the three conditions for the “grouping by similarity” displays and the “low-level groupings” displays are shown in Fig. 1a and b. The “grouping by similarity” displays and the “low-level groupings” displays each accounted for half of the trials, and they were intermixed throughout the experiment. An 8 × 8 array of items was presented in the center of each display. The items were 0.78 cm from their neighbors, both vertically and horizontally, so that the whole array occupied a 6.2 cm × 6.2 cm region.

In the “grouping by similarity” displays, the strips of items could alternate in terms of shape (cross & circle) and color (red & green). As shown in Fig. 1a, the stimuli display of a trial could include color alternation (i.e., color-cue condition), shape alternation (i.e., shape-cue condition), or both alternations (i.e., “both-cues” condition), respectively accounting for one-quarter, one-quarter, and one-half of the trials. In the “low-level groupings” displays, all of the items were solid black dots which could be connected by lines (i.e., connectedness cue) and/or surrounded by rectangles (i.e., common region cue). As shown in Fig. 1b, the stimuli display of a trial could include a connectedness cue (i.e., connectedness condition), a common region cue (i.e., common region condition), or both (i.e., “both-cues” condition), respectively accounting for one-quarter, one-quarter, and one-half of the trials. The directions of the grouping cues were randomized (vertical vs. horizontal), with the constraint that in the “both-cues” condition, the directions of the two grouping cues were always perpendicular to each other. In other words, color and shape, or connectedness and common region, as grouping cues were always in conflict with each other.

The red/green difference in the “grouping by similarity” displays was adjusted for individual participants so that the color and shape-grouping cues were approximately equally strong. This adjustment was implemented by a staircase in the first block: the color difference increased by 10 % if a shape

grouping was chosen in a “both cues” trial but decreased by 10 % if a color grouping was chosen. The contrast of the connection line in the “low-level groupings” displays was adjusted in a similar way so that the connectedness and common region grouping cues were approximately equally strong.

### Procedure

A trial started with a black fixation cross. The fixation cross was presented in the center of the display for 400 ms and was then followed by a gap of 400 ms, after which the stimuli display was presented along with the five response buttons. The stimuli display disappeared after 400 ms, whereas the response buttons remained on the screen until a response was made. This brief exposure was adopted to prevent the participants from swapping back and forth between dimensions. The participants were asked to decide whether, according to their subjective impression, they felt that the structure of the stimuli was (1) clearly horizontal, (2) slightly leaning toward horizontal, (3) approximately balanced without a grouping structure in either direction, (4) slightly leaning toward vertical, or (5) clearly vertical, and then to click the corresponding button (Fig. 1c). The intermediate options were included so that the observers could report the subtle distinctions in their subjective impressions.

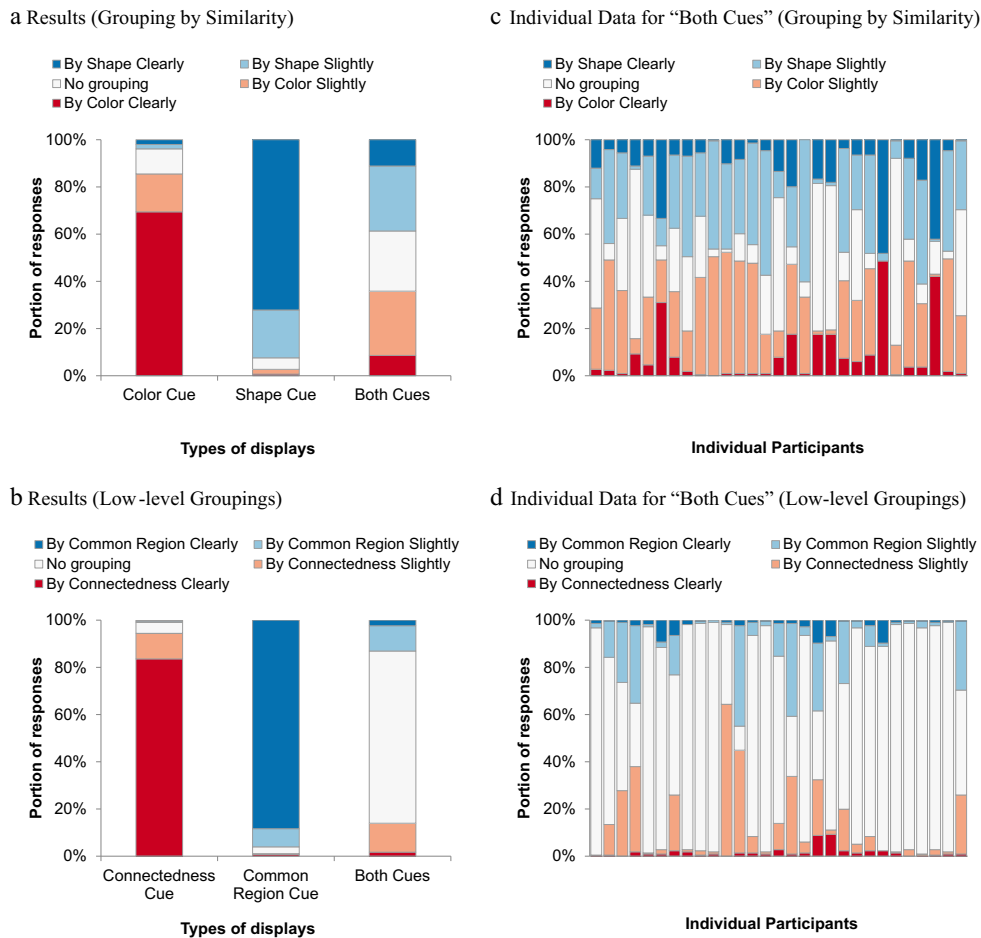
Each participant completed ten blocks (96 trials per block). The first block was regarded as practice and excluded from the analysis.

### Results

In the grouping by similarity displays, the participants’ responses were divided according to whether they conformed to the color-grouping cue, conformed to the shape-grouping cue, or provided “no-grouping” responses (Fig. 2a). When there was only a color-grouping cue in a trial, a “by shape” response indicated grouping reported in the direction opposite to the color grouping, and vice versa.

In the low-level groupings displays, the participants’ responses were divided according to whether they conformed to the connectedness grouping cue, conformed to the common region grouping cue, or provided “no-grouping” responses (Fig. 2b). When there was only a connectedness grouping cue in a trial, a “by common region” response indicated grouping reported in the direction opposite to the connectedness grouping, and vice versa.

As shown in Fig. 2a and b, in both the grouping by similarity displays and the low-level groupings displays, for the conditions in which there was only one unambiguous cue (i.e., color cue, shape cue, connectedness cue, common region cue), the responses were always very consistent with that



**Fig. 2** Results. **a** and **b** show the results from the grouping by similarity displays and the low-level groupings displays, respectively. Most importantly, in the “both-cues” condition of the grouping by similarity displays, the responses were determined by either the color cue or the shape cue, but the “no-grouping” option was not used very frequently (25.5 %). In

contrast, in the “both-cues” condition of the low-level groupings displays, the responses were mainly the “no-grouping” option (73.0 %). **c** and **d** show the individual participant data of the “both-cues” condition in the grouping by similarity displays and the low-level groupings displays, respectively

cue. These high accuracies ensure that the observers were capable of performing the task as instructed.

Most importantly for the present purpose, in the “both-cues” condition of the grouping by similarity displays, the participants often reported grouping in one or the other direction but the “no-grouping” option was not used very frequently (25.5 %). In contrast, in the “both-cues” condition of the low-level groupings displays, the “no-grouping” option was frequently used in responses (73.0 %). The use of the “no-grouping” option was substantially higher in the latter than in the former ( $t(28) = 6.68, p < 0.0001$ ).

To sum up, the present results show that in grouping by similarity, when both color and shape cues are presented and are in conflict with each other in a display, the grouping structure is perceived in a bistable manner: observers perceive either the color cue or the shape cue in individual trials but only occasionally report “no grouping” as their “sum.” However, in low-level groupings, two conflicting cues frequently lead to the impression of “no grouping.”

**Discussion**

The term “grouping” has been used broadly to refer to any process that may organize the visual stimuli from elements into a group. Contour grouping has been shown to operate early in visual processing (Roelfsema 2006; see also Lamme and Roelfsema 2000; Roelfsema, Lamme, and Spekreijse 1998). More broadly, other unambiguous low-level grouping cues include proximity, connectedness, and common region (Franconeri, Bemis, and Alvarez 2009; Palmer and Rock 1994; Rensink and Enns 1995; Trick and Enns 1997; Palmer 1992). However, the present study, along with Levinthal and Franconeri (2011), suggests that grouping by similarity is a product of selection by feature: the same-feature elements become a perceptual unit because they are selected together by the attentional focus. The present study shows that this distinction between grouping by similarity and preattentive groupings also manifests itself in the combination of grouping cues. The

results of the grouping by similarity displays show that similarity cues tend to compete with each other rather than combine into one sum, whereas the results of the low-level grouping displays show that connectedness and common region cues do sum together and result in the impression of no grouping. Together, these results offer strong support for the notion that grouping by similarity is mediated by feature selection (Huang and Pashler 2007; Levinthal and Franconeri 2011).

#### Individual participant analysis

One could suggest that the bistable pattern of responses in the “both-cues” condition of grouping by similarity displays may be caused by a split between two types of participants. If half of the participants always respond to the shape cue and half of the participants always respond to the color cue, then, on average, the result will look like a bistable pattern. To show that this is not the case, the individual-level data of the “both-cues” condition of the grouping by similarity displays are shown in Fig. 2c and those of the “both-cues” condition of the low-level groupings displays are shown in Fig. 2d for comparison. Clearly, there were fairly general bistable patterns in the individual-level data of the grouping by similarity displays (Fig. 2c) but not in the low-level groupings displays (Fig. 2d).

This “split between participants” possibility can also be assessed statistically in the following way: if the participants were merely very biased toward the color or shape cue in the grouping by similarity displays, but their responses to these displays were no more bistable than their responses to the low-level grouping displays, then the portion of the responses to the “less preferred cue for individual participants” should actually be lower in the grouping by similarity displays than in the low-level grouping displays. Such an analysis showed that this portion in the grouping by similarity displays was still substantially higher than that in the low-level grouping displays (33.4 % vs. 11.2 %;  $t(28) = 6.39$ ,  $p < 0.0001$ ), ruling out the “split between participants” possibility.

#### Feature-based attention

In the statement that grouping by similarity has to be made available by feature-based selection, the term “attention” refers not only to the deliberate selection of visual stimuli but also to the situations in which attention can be spontaneously driven by the bottom-up sensory signals of stimuli. For example, when facing the “both-cues” pattern in Fig. 1a, attention may be spontaneously drawn to the red items or to the circles without any explicit intention to do so. This potentially explains why some previous studies have argued that grouping by similarity could occur outside the attentional focus (e.g., Moore and Egeth 1997), which seems to be at odds with the

feature-selection account of grouping by similarity. The grouping structure in Moore and Egeth’s (1997) study was fairly simple, so it remains very possible that the “residual” attention to side stimuli can allow grouping to occur. More decisive evidence for “grouping without attention” needs to be obtained by presenting a large set of items in various features in the periphery and showing that these feature groups are simultaneously formed. To the best of my knowledge, such evidence has not been produced.

#### “Grouping” at the unconscious level

In the feature-selection account of grouping by similarity, I intend to suggest that, at one time, grouping can only occur on one cue by the attentional selection of that feature. However, when a feature is selected by attention, the visual system probably constructs some representations for the items of the other features (i.e., grouping cues). Specifically, Huang and Pashler (2007) described these as the feature-location routines: a mechanism that takes as an input a featural value and returns a Boolean map describing all the locations at which that feature value is present. Therefore, if one regards these “feature-location routines” as “preattentive groups”, then these preattentive groups are probably constructed in parallel. For example, in Fig. 1a, the preattentive groups of the four existing features (i.e., color: red and green; shape: circle and cross) are probably constructed in parallel at the unconscious level, but at one time, only one of these preattentive groups can reach consciousness, control attention, and determine the actual perceived grouping structure.

#### Preattentive combination of low-level grouping cues

With regard to the preattentive combination of low-level grouping cues, two important issues need to be addressed.

First, in Kubovy and Wagemans (1995), bistability has been observed for proximity cues. Why is this different from the present result on low-level grouping cues? Perhaps the reason lies in the nature of the task. In Kubovy and Wagemans (1995), the observers had not been given the option of “no grouping,” and were forced to choose among different grouping structures. Therefore, it seems plausible that, in a considerable portion of trials, these observers had randomly chosen one of these grouping structures even without subjective impression of any grouping structure. This needs to be tested in future studies.

Second, I need to elaborate on the preattentive mechanism that counts all the “votes” of the different low-level grouping cues and provides an overall sum. This mechanism could potentially be implemented in different ways. For example, perhaps there is a “voting machine” for each visual element which counts the votes from all the grouping cues that affect this element, and the overall grouping inclination of this

element will be determined by the “voting result.” Such a machine allows the conflicting cues to cancel each other out, resulting in “zero.” If, for example, an element receives some votes to join a group with its vertical neighbors because of connectedness and an equal number of votes to join a group with its horizontal neighbors because of common region, then the final “voting results” will be “grouping with neither.” Certainly, these speculations on the details of this mechanism need to be tested in future studies and may very well be proven wrong. For the present purpose, what is critical is that this mechanism finishes its computations in early vision preattentively and unconsciously, and reaches a fairly stable perceptual structure and therefore will not lead to bistable percepts.

### Alternative accounts

The critical prediction from the feature-selection account of grouping by similarity, namely that two conflicting similarity grouping cues should lead to bistable percepts, is actually similar to the probabilistic model proposed by Kubovy and van den Berg (2008). Kubovy and van den Berg (2008; see also Kubovy and Wagemans, 1995) suggested that multiple conflicting grouping cues engage in a “competition” for the control of grouping structure and lead to a bistable impression of grouping structures.

The feature-selection account agrees with this general notion of “competition” but further specifies that it is the competition to control the feature-based attention. Clearly, this further specification is necessary. For other types of competitions, there would be no reason to predict the bistable structures in only the combination of similarity grouping cues, but not in the combination of low-level grouping cues.

Kubovy and van den Berg (2008) mentioned two reasons to doubt the feature-selection account of similar grouping (as presented in Huang and Pashler 2007): (1) the effects of attention should have been minimized in brief displays; (2) grouping depends on the feature differences between items, so it is not merely the selection of a single feature. As regards the first reason, attentional selection is known to be very fast and should have sufficient time to show its effect in typical grouping experiments. For example, Huang (2010) showed that the locations of cued items are reported substantially better than those of uncued items in displays as brief as 50 ms. As for the second reason, the efficiency of feature-based selection naturally depends on the target/distractor differences. For example, the selection of red among pinkish-red items will certainly be more difficult than the selection of red among green items. So the effect of the feature difference throws no doubt on the feature selection account.

Another potential alternative account is that the grouping cues are always combined preattentively. However, this account would then have trouble explaining the bistable reports for conflicting similarity cues.

All in all, the feature-selection account of grouping by similarity offers the best explanation of the results. The other alternative accounts illustrated above (i.e., general competition between the grouping cues, or general preattentive combination) cannot explain the distinction between grouping by similarity and low-level groupings.

**Acknowledgments** This research was supported by RGC Grant 446411 from the Hong Kong government.

### References

- Attneave, F. (1971). Multistability in perception. *Scientific American*, 225, 63–71.
- Franconeri, S. L., Bemis, D. K., & Alvarez, G. A. (2009). Number estimation relies on a set of segmented objects. *Cognition*, 113, 1–13.
- Huang, L. (2010). The speed of feature-based attention: Attentional advantage is slow but selection is fast. *Journal of Experimental Psychology-Human Perception and Performance*, 36, 1382–1390.
- Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review*, 114, 599–631.
- Kubovy, M., & Van Den Berg, M. (2008). The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. *Psychological Review*, 115, 131.
- Kubovy, M., & Wagemans, J. (1995). Grouping by proximity and multistability in dot lattices: A quantitative Gestalt theory. *Psychological Science*, 6, 225–234.
- Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 23, 571–579.
- Levinthal, B. R., & Franconeri, S. L. (2011). Common-fate grouping as feature selection. *Psychological Science*, 22, 1132–1137.
- Moore, C. M., & Egeth, H. (1997). Perception without attention: Evidence of grouping under conditions of inattention. *Journal of Experimental Psychology-Human Perception and Performance*, 23, 339–352.
- Palmer, S. E. (1992). Common region: A new principle of perceptual grouping. *Cognitive Psychology*, 24, 436–447.
- Palmer, S., & Rock, I. (1994). Rethinking perceptual organization - the role of uniform connectedness. *Psychonomic Bulletin & Review*, 1, 29–55.
- Rensink, R. A., & Enns, J. T. (1995). Preemption effects in visual-search - evidence for low-level grouping. *Psychological Review*, 102, 101–130.
- Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience*, 29, 203–227.
- Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, 395, 376–381.
- Trick, L. M., & Enns, J. T. (1997). Clusters precede shapes in perceptual organization. *Psychological Science*, 8, 124–129.