

Short article

A conjunctive feature similarity effect for visual search

Yuji Takeda, Steven Phillips, & Takatsune Kumada

National Institute of Advanced Industrial Science and Technology (AIST), Ibaraki, Japan

Traditional models of visual search assume interitem similarity effects arise from within each feature dimension independently of other dimensions. In the present study, we examine whether distractor–distractor effects also depend on feature conjunctions (i.e., whether feature conjunctions form a separate “feature” dimension that influences interitem similarity). Spatial frequency and orientation feature dimensions were used to generate distractors. In the bound condition, the number of distractors sharing the same conjunction of features was higher than that in the unbound condition, but the sharing of features within frequency and orientation dimensions was the same across conditions. The results showed that the target was found more efficiently in the bound than in the unbound condition, indicating that distractor–distractor similarity is also influenced by conjunctive representations.

Since the early 1980s, the nature of attention in visual search has been extensively investigated by a large number of researchers. It is generally accepted that search efficiency is affected by two types of similarity: target–distractor similarity and distractor–distractor similarity. Wolfe, Cave, and Franzel (1989) demonstrated that a target could be found more efficiently when the target shared only one feature with distractors than when two features were shared in a triple conjunction search task. This is the effect of the target–distractor similarity. On the other hand, the effect of the distractor–distractor similarity was reported by Duncan and Humphreys (1989). In their study, observers were required to look for a T among Ls. They found that search performance improved

when distractors were displayed with the same orientation compared to when they were rotated randomly. In summary, search efficiency increased with decreasing target–distractor similarity and with increasing distractor–distractor similarity.

Guided search model (GSM) is a prominent model of visual search, which can explain both the effects of target–distractor similarity and distractor–distractor similarity on search efficiency (Cave & Wolfe, 1990; Wolfe, 1994; Wolfe et al., 1989). GSM assumed that the visual input is decomposed through several feature dimensions, such as colour, orientation, or motion, and that target–distractor and distractor–distractor similarity are calculated in each feature dimension separately. An activation map, which guides

Correspondence should be addressed to Y. Takeda, Cognition and Action Group, Institute for Human Science and Biomedical Engineering, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 6, 1–1–1 Higashi, Tsukuba City, 305–8566 Ibaraki, Japan. E-mail: yuji-takeda@aist.go.jp

observer's attention, is composed of the linear summation of similarities in all feature dimensions. A key concept relevant to the present study is that similarities are defined in each feature dimension. This concept is succeeded in many recent models of visual search (e.g., Itti & Koch, 2000; Nothdurft, 2000). Although the findings in neurophysiological studies support the account that features are handled by separate modules in early visual processing (Zeki, 2001, for review), there is no evidence to suggest that similarities are calculated only by such modules with regard to visual search. Rather, some studies have proposed that combinations of two feature dimensions, such as motion and form (McLeod, Driver, & Crisp, 1998), or binocular disparity and colour (Nakayama & Silverman, 1986), are processed preattentively, in a way that affects search efficiency. These findings suggest the possibility that similarity effects in visual search are not only influenced by the relationships between display item features within each feature dimension, but also by the relationships between display item feature conjunctions.

In the present study, we examined whether distractor–distractor similarity in visual search also depends on conjunctive features in a way that cannot be decomposed into a linear combination of feature components, as generally assumed in models of visual search. Feature dimensions of spatial frequency and orientation were used to construct distractor items. Each feature dimension had three levels, and all feature levels appeared with equal probability. Two types of distractor sets were produced by manipulating feature combinations. One distractor set consisted of nine variations (3 spatial frequencies \times 3 orientations). The other distractor set consisted of three variations (three levels of spatial frequency and three levels of orientation were used, but their combination was fixed). In this case, GSM, as with many other models, predicts no difference in search efficiency for the two types of distractor set, because the variation of features in each dimension is the same. On the other hand, if distractor–distractor similarity is also based on the representation of bound features,

distractor–distractor similarity and therefore search efficiency would be greater in the latter condition.

Method

Participants

A total of 8 graduate or undergraduate students with normal or corrected-to-normal vision participated as paid volunteers in this experiment.

Apparatus and stimuli

Stimuli were displayed on a 17-in. colour monitor (IIYAMA A702, with $1,024 \times 768$ -pixel resolution) controlled by an IBM PC AT-compatible computer (EPSON Pro2500).

Gabor patches, which subtended 2.0° of visual angle, were used as search items. The spatial frequency of these items was 1, 2, 3, or 4 cycle/deg, and the orientation was 0° , 45° , 90° , or 135° . A total of 10 or 19 items were presented at one of 64 possible locations (an 8×8 invisible matrix subtended $20.6^\circ \times 20.6^\circ$) with 0.5° of jitter (see Figure 1). All stimuli were presented on a grey background.

Design

There were three main variables in this experiment. The first variable was target-present versus target-absent trials. In 50% of trials, a target was presented in the display. The target was a Gabor patch, with a frequency of 3 cycles/deg and an orientation of 135° (tilted 45° to the right of vertical). The second variable was search set size. Set sizes of 10 and 19 were used with equal probability. The third variable was distractor–distractor similarity. The distractors were Gabor patches, with a frequency of 1, 2, or 4 cycles/deg and an orientation of 0° , 45° , or 90° . In the bound condition, a fixed combination of frequency and orientation features was used (see Figure 1). In the unbound condition, the distractor set consisted of all combinations of frequency and orientation—namely, nine types of distractor appeared in the display. Conditions were balanced and randomly ordered within each block.

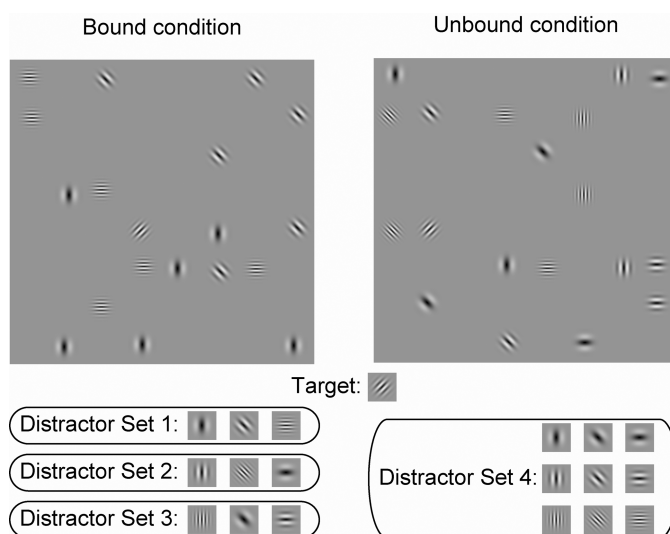


Figure 1. Example search stimuli for bound (left) and unbound (right) conditions. The target was a Gabor patch, with a frequency of 3 cycles/deg and an orientation of 135 degrees. Distractor Sets 1, 2, or 3 were used in the bound condition, and Distractor Set 4 was used in the unbound condition.

Procedure

The participants observed the stimuli at a viewing distance of 58 cm in a semidark room. Each trial began with the presentation of a fixation cross (“+”, $0.32^\circ \times 0.32^\circ$) at the centre of the display. After 1,000 ms, the fixation cross was removed, and the search display was presented. Participants were required to respond as quickly and accurately as possible to indicate whether or not the display contained the target by pressing one of two keys with their right or left index finger. All stimuli vanished after the participant’s response, and there was a 667-ms blank period between trials. No feedback was given to the participants.

Each participant performed 1 practice block and 24 experimental blocks. Each block consisted of 48 trials: 2 (target-present and target-absent trials) \times 2 (set sizes of 10 and 19) \times 2 (bound and unbound conditions) \times 6 repetition.

Results

Incorrect response trials, or trials with reaction times (RTs) longer than three standard deviations

from the mean of each condition (1.6%), were excluded from RT analysis. The mean correct RTs and error rates as a function of set size are shown in Figure 2. The search slopes (i.e., the RT/set size gradients) were 18.2 ms/item (bound) and 27.4 ms/item (unbound) for target-present trials and 59.5 ms/item (bound) and 76.2 ms/item (unbound) for target-absent trials. These slopes indicate that participants could not find a target efficiently in the present search task.

RTs were subjected to a three-way analysis of variance (ANOVA) with target (present vs. absent), set size (10 vs. 19), and similarity (bound vs. unbound) as the main terms. All main effects were significant, $F(1, 7) = 75.7$, $p < .0001$, for target; $F(1, 7) = 157.4$, $p < .0001$, for set size; $F(1, 7) = 124.9$, $p < .0001$, for similarity. In addition, the two-way interactions of Target \times Set Size, $F(1, 7) = 36.5$, $p < .001$, Target \times Similarity, $F(1, 7) = 53.1$, $p < .001$, and Set Size \times Similarity, $F(1, 7) = 52.4$, $p < .001$, were significant, but the three-way interaction was not significant, $F(1, 7) = 0.69$, $p = .43$. The results indicate that search in the bound condition was easier than that in the unbound

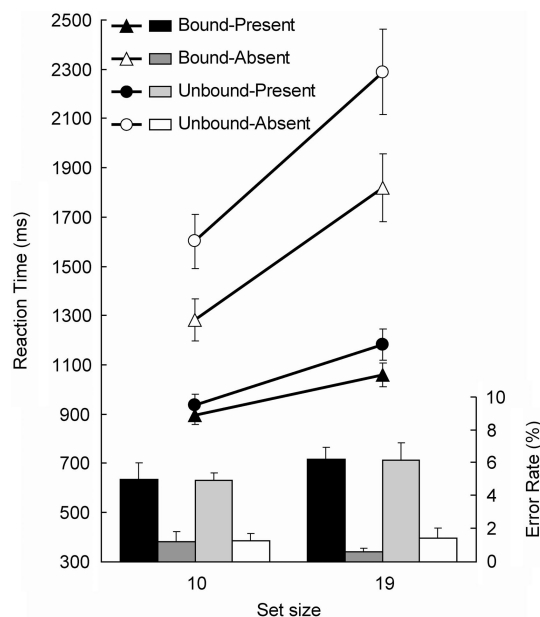


Figure 2. Mean reaction time (ms) as a function of set size for target-present (filled) and target-absent (open) trials in the bound (triangle) and unbound (circle) conditions. The histogram shows error rates for target-present bound (black), target-absent bound (dark grey), target-absent bound (light grey), and target-absent unbound (white) trials. Error bars in both graphs indicate 1 standard error.

condition, and this advantage for the bound condition increased with set size. Furthermore, the advantage was larger in the target-absent trials than in the target-present trials.

Error rates were also subjected to a three-way ANOVA with the same factors as those in the RT analysis. Only the main effect of target was significant, $F(1, 7) = 235.9, p < .0001$. This indicates that error rates were larger in the target-present trials than in the target-absent trials.

Discussion

The results show that interitem similarity is also influenced by second-order (i.e., conjunctive) features. The target was found more efficiently in the bound than in the unbound condition. Yet, the number of distractors sharing features with other distractors along each dimension was the same

for each condition. That is, each distractor item shared its frequency (or orientation) feature with 2 other distractors when 9 distractors were presented and with 5 other distractors when 18 distractors were presented. Therefore, the advantage for the bound condition is not explained on the basis of a linear combination of dimension-specific feature similarities—similarity over first-order feature maps, as generally assumed in visual search models.

Our result has implications for theories/models of visual search that rely on perceptual grouping to explain distractor–distractor similarity effects in visual search. Like attention engagement theory (Duncan & Humphreys, 1989), we acknowledge that distractor–distractor similarity causes attention to these items to be suppressed. Computationally, distractors sharing the same features may be grouped and rejected at once (Grossberg, Mingolla, & Ross, 1994). However, in contrast to these and related explanations (Bundesen, 1998; Logan, 1996; Wolfe et al. 1989), we suggest that conjunctions of features form an additional dimension that contributes to interitem similarity. Previous work has focused on the effects of first-order similarity by manipulating the sharing of distractor features along individual dimensions (e.g., orientation) without independently manipulating the sharing of feature conjunctions (e.g., colour–orientation). First-order similarity cannot account for the conjunctive feature effect observed here, because it was held constant across experiment and baseline conditions. Models that traditionally rely on first-order feature maps only (e.g., GSM) or those that do not consider second-order grouping (e.g., Duncan & Humphreys, 1989; Grossberg et al., 1994) would need to be extended, possibly by treating conjunctive features as an additional dimension, to address this sort of data.

A model based on concurrent feature and conjunctive feature maps is in stark contrast to stage models, where features and conjunctions have distinct computational roles, such as those derived from feature integration theory (Treisman & Gelade, 1980). Yet, recent studies in visual

categorization lend some support to the existence of early (i.e., preattentive) feature binding processes (e.g., Li, VanRullen, Koch, & Perona, 2002; Rousselet, Fabre-Thorpe, & Thorpe, 2002). In these studies, observers determined whether a natural scene contained an instance of a target category (e.g., animal). Rapid categorization may just be afforded by well-trained neurons that fire only when a stimulus lies within a specific region of feature space (cf. feed-forward connectionist networks; McClelland, Rumelhart, & Group, 1986). However, such processes may not be involved here, because they would be insensitive to the relationships between distractor stimuli, which lie outside the region that defines the target. Conjunctive feature similarity adds another twist to an account of visual search efficiency, although further work is needed to ascertain exactly what role conjunctive features play with respect to this effect.

Original manuscript received 25 July 2006
Accepted revision received 11 October 2006
First published online day month year

REFERENCES

- Q2 Bundesen, C. (1998). A computational theory of visual attention. *Philosophical Transactions of the Royal Society of London, Series B*, 353, 1271–1281.
- Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology*, 22, 225–271.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96, 433–458.
- Grossberg, S., Mingolla, E., & Ross, W. D. (1994). A neural theory of attentive visual search: Interactions of boundary, surface, spatial, and object representations. *Psychological Review*, 101, 470–489.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40, 1489–1506.
- Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, 99, 9596–9601.
- Logan, G. D. (1996). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review*, 103, 603–649.
- McClelland, J., Rumelhart, D., & Group, P. R. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vols. I & II). Cambridge, MA: MIT Press.
- McLeod, P., Driver, J., & Crisp, J. (1988). Visual search for a conjunction of movement and form is parallel. *Nature*, 332, 154–155. Q1
- Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, 320, 264–265.
- Nothdurft, H. C. (2000). Saliency from feature contrast: Additivity across dimensions. *Vision Research*, 40, 1183–1201.
- Rousselet, G., Fabre-Thorpe, M., & Thorpe, S. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience*, 5, 629–630.
- Treisman, A. M., & Gelade, G. (1980). A feature integration theory of perception. *Cognitive Psychology*, 12, 97–136.
- Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1, 202–238.
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception & Performance*, 15, 419–433.
- Zeki, S. (2001). Localization and globalization in conscious vision. *Annual Review of Neuroscience*, 24, 57–86.

PQJE206220

Queries

Yuji Takeda, Steven Phillips, & Takatsune Kumada

Q1 McLeod, Driver, & Crisp: 1988 in refs, 1998 in text.

Q2 Bundesen (1998): Bundesen in refs, Bundasen in text.