A surface-based code contributes to visual shape perception

Martin Arguin
Centre de Recherche en Neuropsychologie Expérimentale et Cognition, Département de psychologie, Université de Montréal, Montréal, Canada

Ian Marleau
CISSS de la Montérégie-Ouest, Installation Longueuil, Longueuil, Canada

Mercédès Aubin
CÉGEP de Jonquière, Département des Sciences humaines, Jonquière, Canada

Sacha Zahabi
Centre de Recherche en Neuropsychologie Expérimentale et Cognition, Département de psychologie, Université de Montréal, Montréal, Canada

E. Charles Leek
School of Psychology, Institute of Life and Human Sciences, University of Liverpool, Liverpool, UK

Considerable uncertainty remains regarding the types of features human vision uses for shape representation. Visual-search experiments are reported which assessed the hypothesis of a surface-based (i.e., edge-bounded polygons) code for shape representation in human vision. The results indicate slower search rates and/or longer response times when the target shape shares its constituent surfaces with distractors (conjunction condition) than when the target surfaces are unique in the display (nonconjunction condition). This demonstration is made using test conditions that strictly control any potential artifact pertaining to target–distractor similarity. The surface-based code suggested by this surface-conjunction effect is strictly 2-D, since the effect occurs even when the surfaces are shared between the target and distractors in the 2-D image but not in their 3-D instantiation. Congruently, this latter finding is unaltered by manipulations of the richness of the depth information offered by the stimuli. It is proposed that human vision uses a 2-D surface-based code for shape representation which, considering other key findings in the field, probably coexists with an alternative representation mode based on a type of structural description that can integrate information pertaining to the 3-D aspect of shapes.

Introduction

Shape is probably the most important visual property of an object with respect to our capacity to recognize and interact with it. Fundamental to our understanding of shape perception is knowledge of the code(s)—that is, the type(s) of property—by which the human visual system represents this property. Candidate codes that have been proposed in the literature are very diverse and cover a vast range: from local, low-level properties such as oriented edges or vertices (e.g., Edelman & Weinshall, 1991; Lades et al., 1993; Poggio & Edelman, 1990; Ullman, 1989) and Fourier descriptors (e.g., Cortese & Dyre, 1996), to intermediate-level properties such as surface fragments (Ullman, 2007), aspect ratio or tapering (Arguin, Bub, & Dudek, 1996; Arguin & Saumier, 2000; Saumier & Arguin, 2003; Stankiewicz, 2002), and shape constraints—that is, "a priori principles involving spatially global aspects of 3D objects such as symmetry or compactness" (Pizlo, 2008, p. 183), to complex high-order three-dimensional features such as generalized cylinders (Brooks, 1981; Marr, 1982; Marr & Nishihara, 1978), geons (Biederman, 1987; Hummel & Biederman, 1992), superquadrics (Barr, 1981; Pentland, 1986), and medial axis

Citation: Arguin, M., Marleau, I., Aubin, M., Zahabi, S., & Leek, E. C. (2019). A surface-based code contributes to visual shape perception. *Journal of Vision*, *19*(11):6, 1–23, https://doi.org/10.1167/19.11.6.

https://doi.org/10.1167/19.11.6 Received July 17, 2018; published September 11, 2019 ISSN 1534-7362 Copyright 2019 The Authors

structure (Blum, 1973; Burbeck & Pizer, 1995; Feldman & Singh, 2006; Hung, Carlson, & Connor, 2012; Kimia, 2003; Marr & Nishihara, 1978).

The nature of the codes actually underlying human visual shape perception remains a matter of debate. There is evidence, however, that features of intermediate complexity may be optimal for object classification (Crouzet & Serre, 2011; Delorme, Richard, & Fabre-Thorpe, 2010; Nakayama, He & Shimojo, 1995; Nakayama & Shimojo, 1992; Ullman, Vidal-Naquet, & Sali, 2002). Within that broad class of features, the use of object surfaces as the basis for shape description has historically proven very effective in machine vision (e.g., Ashbrook, Fisher, Robertson, & Werghi, 1998; Barrow & Tenenbaum, 1981; Fan, Medioni, & Nevatia, 1989; Faugeras et al., 1983; Fisher, 1989; Lee & Park, 2002; Potmesil, 1983).

The purpose of the present research is to assess the contribution to visual shape perception of an intermediate-level code based on the collection of surfaces (defined here as edge-bounded polygons) that determine the shape of an object (Leek, Reppa, & Arguin, 2005; Leek, Reppa, Rodriguez, & Arguin, 2008; Leek, Roberts, Dundon, & Pegna, 2018; Reppa, Greville, & Leek, 2015; Reppa & Leek, 2019). This hypothesis assumes that the contours present in the image of an object lead to a segmentation of the stimulus into its visible surfaces, the collection of which constitutes its online perceptual description. For recognition, this description may be matched to a similarly structured surface-based memory representation which may include currently unexposed surfaces (because of occlusion, for instance), provided that they are familiar to the observer from previous experience. This view is highly congruent with the proposal of Nakayama and Shimojo (1992; see also Nakayama et al., 1995) that the representation of visible surfaces constitutes a crucial stage upon which high-level vision (including shape perception and object recognition) must rest. It also shares similarities with the fragment-based hierarchy approach (Ullman, 2007; Ullman & Bart, 2004; Ullman et al., 2002; see also Ullman, Assif, Fetaya, & Harari, 2016), in which shape is determined by a collection of informative fragments. The fragments extracted during categorization are then used to build the fragment-based memory representation for recognition.

Empirical evidence in support of surface-based shape perception has been reported previously using a whole-part matching paradigm with line drawings as stimuli (Leek et al., 2005; for similar and related findings, see also Reppa et al., 2015; Reppa & Leek, 2019). The results indicate an equal and significant performance advantage when part stimuli are volumetric components (e.g., geons) or closed surfaces relative to when they are either open or closed contours that do not form a complete object surface. Significantly, these effects cannot be attributed to a mismatch across conditions on low-level stimulus properties of the part stimuli (e.g., length of edge contour, number of edge vertices, or number of visible surfaces), nor to image overlap between the part and the object to be matched, which were both controlled. Since the most parsimonious description of the two stimulus conditions that led to equal best performance is in terms of non-volumetric edge-bounded regions (i.e., surfaces), this was taken to reflect the representational code mediating the performances observed.

A subsequent study by Leek et al. (2008) used a priming paradigm within the context of an object-recognition (old/new) task. The findings from one experiment indicated that inferred surfaces (i.e., those not visible, due to self-occlusion, but which should be part of the perceptual description of the object assuming a code based on geometric volumes) do not prime object recognition, whereas visible surfaces do. Moreover, priming from visible surfaces showed no advantage according to whether those surfaces constituted a volumetric component or not (i.e., belonged to distinct components). The most parsimonious account of the results is that a surface-based code mediated performance. In contrast, the findings contradict the notion of volumetric completion that is required by the hypothesis of a representation based on volumetric components.

In the present study, we assess the hypothesis of a surface-based shape representation, using as diagnostic the surface-conjunction effect in the visual-search task. The conjunction effect has been frequently used in the past to investigate a range of issues in human vision, notably to determine the features used by the visual system to code stimuli (e.g., Treisman & Gelade, 1980; Treisman & Gormican, 1988; Treisman & Sato, 1990). The underlying principle is relatively simple. When a target is made of a combination of features that it shares with distractors, visual search will be slower than when features are not shared with distractors if these features correspond to the code involved in the representation of the target. For instance, if the target is a green vertical bar and distractors are green horizontals and red verticals, the search rate will be slower than if all distractors are either green horizontals or red verticals. Such a finding is typically taken to indicate that color and orientation are basic properties that subtend the representation of visual stimuli.

This fundamental observation by Treisman and colleagues, that the processing of feature conjunctions poses a particular challenge for the human visual system, has led to the formulation of Feature Integration Theory (Treisman & Gelade, 1980), which suggests that the performance cost associated with the integration of visual features must be attributed to the unique capacity of spatially focused attention in this respect.

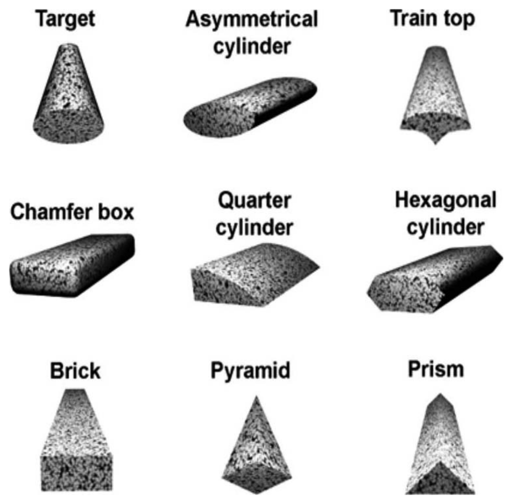

Figure 1. Objects used as stimuli in Experiment 1.

This interpretation has been challenged by several authors on various empirical and theoretical grounds, with the alternative view proposed being that attention is not required for feature integration and that the performance cost in processing feature conjunctions may be explained otherwise (e.g., Carrasco, Evert, Chang, & Katz, 1995; Carrasco, Giordano, & McElree, 2006; Carrasco & Yeshurun, 1998; Dugué, Xue, & Carrasco, 2017; Duncan & Humphreys, 1989, 1992; Eckstein, 2001; McElree & Carrasco, 1999; Palmer, Verghese, & Pavel, 2000). In the present study, we remain neutral with respect to the specific nature of the process required to encode conjunctions of visual features, while adopting the consensus view that "conjunction searches are … more complex than feature searches, as they require combining information from two [or more] separate feature dimensions" (Carrasco et al., 2006, p. 2029).

Use of the feature-conjunction effect to understand human vision has been adapted to investigate the contribution of global shape properties to visual shape representation (Arguin & Saumier, 2000; Saumier & Arguin, 2003). For example, the target may be a filled 2-D pattern looking like the silhouette of a banana, whereas one set of distractors is made of shapes which have the same aspect ratio as the target but are straight, and the other set of distractors is made of shapes with a curvature similar to the target's but a different aspect ratio. The rate of visual search for such a conjunction target is slower than that for a target that does not share its constituent features with distractors, even though the conjunction and nonconjunction conditions are matched with respect to target/distractor similarity (for further details, see Arguin & Saumier, 2000; Saumier & Arguin, 2003; for an extension of the approach to investigate the effects of 2-D and 3-D rotations, see also Blais, Arguin, & Marleau, 2009).

Here, the role of surfaces in the representation of visual shape is examined using targets that either share or do not share their constituent surfaces with the distractors displayed with them (Figure 1). The hypothesis that a surface-based representation contributes to shape perception predicts that visual search will be slower when the target is made of a conjunction of distractor surfaces (conjunction condition) than when it is not (nonconjunction condition).

Beyond providing an empirical test for a surface-based code for shape representations, the present research also aims to characterize the surface code with respect to its capacity to integrate depth information—that is, to represent shape properties in depth rather than in 2-D only. Shape-perception theories diverge on this issue, even among those that propose a shape representation based upon a volumetric code.

For instance, the computational theory of Marr (1982; Marr & Nishihara, 1978) proposes that shape perception integrates depth information at its 2.5-D stage, which represents the depth orientation of visible surfaces. In contrast, the recognition-by-components theory assumes that geons (i.e., the volumetric primitives it proposes) are inferred from 2-D contour information only (Biederman, 1987). Likewise, Pizlo's (2008) theory based on shape constraints argues that depth information should fail to contribute to the perception of 3-D shape. Invariant recognition for different viewing conditions may also be obtained without using any 3-D information (Crouzet & Serre, 2011; Ullman, 2007). As for the notion of a surface-based code for shape perception (Leek et al., 2005; Leek et al., 2008; Reppa et al., 2015) considered here, its current formulation assumes that the surfaces of an object are coded in terms of 2-D polygons, without any representation of depth information. An alternative possibility, however, is that the surface code may represent the depth orientation of surfaces relative to the observer, as determined from depth information such as shading, texture, or binocular disparity, for instance (Nefs, 2008; Saunders & Backus, 2006; Todd, Norman, Koenderink, & Kappers, 1997).

On the empirical front, even though human vision is limited in its capacity to represent surfaces in depth (Norman & Todd, 1996, 1998; Norman, Todd, Norman, Clayton, & McBride, 2006; Norman, Todd, & Philips, 1995; Todd & Norman, 2003), there is evidence indicating that depth information may actually contribute to visual shape perception. For instance, Aubin and Arguin (2014) have reported that shading and stereoscopic depth have a major impact on judgements of the convexity or concavity of an otherwise ambiguous synthetic shape. Burke (2005) has shown that stereoscopic information reduces the cost of depth rotation in the recognition of bent paper-clip shapes (for related findings, see also Bennett & Vuong, 2006).

A stereo-display advantage has also been reported by Cristino, Davitt, Hayward, and Leek (2015; see also Oliver, Cristino, Roberts, Pegna, & Leek, 2018) using objects made of an assembly of geon-like elements, but only when the objects to be matched are shown from markedly different viewpoints. A study by Lee and Saunders (2011) has shown that stereo reduces the cost of depth rotation, but also improves shape discrimination even when rich monocular shape cues are available. Similar observations on the effect of stereoscopic depth information have also been reported for face recognition across changes in viewpoint (Burke, Taubert, & Higman, 2007; Hong Liu, Ward, & Young, 2006).

Another study by Blais et al. (2009) has shown specifically that a code for shape representation based on global properties such as aspect ratio or tapering (Arguin et al., 1996; Arguin & Saumier, 2000; Saumier & Arguin, 2003; Stankiewicz, 2002) is invariant to depth (as well as 2-D) rotations, on the condition that the stimulus offers sufficient depth information. On the other hand, other studies have found that stereo presentation may produce a disadvantage for recognition under particular test conditions (Pasqualotto & Hayward, 2009). Whether this particular evidence extends to a surface-based shape-representation code remains unknown, however, since it may rest on markedly different stimulus information than that driving performance in these studies.

The goal of Experiments 2–4 was to assess jointly whether a surface code contributes to human shape perception and whether it integrates depth information. This was done by examining the surface-conjunction effect using surfaces that are shared between the target and distractors in terms of their 2-D contours but not in their 3-D instantiation. This is illustrated by the stimuli in the top row of Figure 1. The target, which is cone shaped, has a base with the same 2-D contours as the end of the Asymmetrical cylinder distractor. Similarly, the target shares the 2-D contours of its side with the top portion of the Train top distractor. However, it can easily be seen that these so-called shared surfaces are actually not shared in their 3-D instantiation. Thus, whereas the base of the target is perfectly circular in 3-D, the extremity of the asymmetrical cylinder is elliptical. As well, the side of the cone has a 3-D shape that is completely different from the top surface of the Train top.

Under these conditions and with the appropriate distractor set for a baseline nonconjunction condition (determined in Experiment 1), the occurrence of a surface-conjunction effect would suggest disregard of depth information—that is, that the surface-based code rests on 2-D contour information. In contrast, if the surface-based code fully integrates depth information, then the surfaces that it will represent for the target and distractors described in the conjunction condition will not match, and no conjunction effect should occur.

As an additional test for the role of depth information in a surface-based code, Experiments 2–4 reported here used stimuli that vary in the richness of the depth information they offer. Thus, in Experiment 2, the stimuli comprised a rich surface texture, which will add to shading as a cue for the depth orientation of the surfaces of objects. Depth information was further enhanced in Experiment 3, which used the same textured objects as Experiment 2, but this time presented as anaglyphs that offer realistic binocular disparity information. In contrast, in Experiment 4 stimuli had no surface texture and no binocular disparity, thereby limiting the depth information available to shading. Under the conditions already described (see Figure 1), disregard for depth information by a surface-based code would be indicated by a surface-conjunction effect that is invariant to the richness of depth information. In contrast, a modulation of the surface-conjunction effect by depth information would reveal the capacity of the surface code to integrate this information, at least to some degree.

Thus, Experiment 1 used a simultaneous matching task to identify a pair of distractors which do not share any surface with the target shape but which are equated to the conjunction distractors on their similarity with the target. Experiment 2 used the stimuli selected in Experiment 1 to assess the surface-conjunction effect with realistically rendered textured 2-D shapes. Experiments 3 and 4 were designed similarly, but the quality of depth information is enhanced by binocular disparity in Experiment 3, whereas it is degraded in Experiment 4 by the use of textureless 2-D shapes.

Experiment 1

The purpose of Experiment 1 was to select distractors for the nonconjunction condition that are matched to those of the conjunction condition in terms of their similarity to the target. These were then used to assess the surface-conjunction effect in Experiments 2–4, where the target shared the 2-D contours defining its surfaces with the conjunction distractors, but with no shared surface in terms of their 3-D instantiation. Thus, the distractors for the conjunction condition are characterized by the fact that they share the 2-D projection of the contours defining one of their surfaces with the target. The surfaces making up the objects, however, are not shared between items in their 3-D instantiation. The stimuli used in Experiment 1 as candidate distractors for the nonconjunction condition do not share any surface with the target (see second and third rows of Figure 1). The task used is that of simultaneous matching for two shapes aligned vertically, one above and the other below the fixation point. The critical data comes from the "different" trials, where one stimulus was the target and another was a distractor, either from the conjunction condition or one of the candidate nonconjunction distractors. Performances (correct response times and error rates) with individual conjunction distractors were compared to those obtained with individual nonconjunction distractors. The nonconjunction distractors that led to performance not significantly different from, and as close as possible to, that with the conjunction distractors were selected to serve as the nonconjunction distractors in Experiments 2–4.

Method

Participants

Twelve volunteers (four men and eight women) aged between 18 and 29 years took part in the experiment. All were unaware of the purpose of the experiment, were neurologically intact, and had normal or corrected-to-normal visual acuity. No particular selection criteria were applied with respect to gender, manual dominance, or level of education.

Stimuli

The stimuli are shown in Figure 1. They were simple synthetic objects created using the 3D Studio MAX program from Autodesk. A total of nine objects were used. One was the target (Cone) and two others were the distractors for the conjunction condition (Asymmetrical cylinder and Train top). The latter share the 2-D projection of the contours of one of their surfaces with the target. These shared 2-D surfaces, however, are not shared in their 3-D instantiation. Thus, whereas the visible end of the Asymmetrical cylinder has 2-D contours identical to the base of the target, the shared surfaces are quite different in their 3-D instantiation, as signaled by the gradients of their surface textures as well as by shading. The same is true for the side of the target and the top portion of the Train top distractor. The six other objects (Chamfer box, Quarter cylinder, Hexagonal cylinder, Brick, Pyramid, and Prism) were the nonconjunction distractors and shared none of their surfaces (either in 2-D or 3-D) with the target. The depth orientation of the main axis of one subset of these nonconjunction distractors was identical to that of the Asymmetrical cylinder, whereas the other subset shared the depth orientation of its main axis with that of the Train top (see Figure 1). A monochromatic texture was applied on the surface of these objects using 3D Studio MAX. Objects were then rendered as 2-D images, which were the actual stimuli used in the experiment. The spatial extent of the stimuli ranged between 1.3° and 2.6° of visual angle horizontally, and between 1.0° and 1.7° vertically, from the viewing distance of 95 cm.

Procedure

Participants were instructed to indicate as rapidly as possible, while avoiding errors, whether the two items displayed on each trial were the same or different. Two experimental blocks of 240 trials each (total of 480 trials) were conducted, preceded by a practice block of 32 trials. Stimulus pairs were identical on half of the trials and different on the other half, and the order of trials was random.

Each trial began with a fixation cross (+ character in Courier font, 26 points) of 500 ms duration. Its offset was immediately (interstimulus interval = 0 ms) followed by the display of the stimuli, which were aligned vertically and shown at a distance of 1.6° from fixation. The stimuli remained visible until the participant's response. Each of the shapes used in Experiment 1 served on an equal number of occasions in "same" trials. On "different" trials, one of the items was always the cone (to serve as the target in the following visual-search experiments), whereas the other item was the one of the other shapes, each of these serving on an equal number of trials. Responses consisted of pressing the appropriate button of a response box interfaced with the computer controlling the experiment using the left or right index finger for the left and right response buttons, respectively. The use of the left and right response buttons to indicate that stimuli were the same or different was counterbalanced across participants.

Results and discussion

Individual correct response times (RTs) that were more than 2.5 standard deviations from the participant's mean for that condition were excluded from data analysis. This resulted in the elimination of 261 trials (4.5% of all trials). The correlation between correct RTs and error rates was 0.09 (not significant), thus indicating no speed/accuracy trade-off.

The crucial data from Experiment 1 are the correct RTs and error rates on "different" trials, where one distractor was displayed along with the cone designed to serve as the target in Experiment 2. Data analyses thus focused on this set of trials and compared performance between trials involving distractors from the conjunction condition (i.e., Asymmetrical cylinder and Train top) to those with distractors that do not share any surface with the target and that have the same depth orientation of their main axis. Mean RTs

Conjunction distractors	Nonconjunction distractors		
	Chamfer box 541.9 ms (0.022)	Hexagonal cylinder 572.7 ms (0.006)	Quarter cylinder 543.0 ms (0.011)
Asymmetrical cylinder 555.4 ms (0.036)	$F < 1$ $F < 1$	$F < 1$ $F = 4.5$ (not significant)	$F < 1$ $F = 3.1$ (not significant)
	Brick 611.7 ms (0.014)	Pyramid 635.3 ms (0.039)	Prism 595.9 ms (0.014)
Train top 599.2 ms (0.019)	$F < 1$ $F < 1$	$F = 3.4$ (not significant) $F = 1.9$ (not significant)	$F < 1$ $F < 1$

Table 1. Correct response times (and error rates) on "different" trials in Experiment 1 along with the relevant statistical tests (see text for details). All degrees of freedom for the F statistics are (1, 11).

and error rates, as well as the outcomes of the statistical tests, are displayed in Table 1. Given the observations in Table 1, the Chamfer box was chosen as the nonconjunction distractor to serve in the following experiments as a control for the Asymmetrical cylinder conjunction distractor. Indeed, the Chamfer box offers the closest performance to that observed with the Asymmetrical cylinder when RTs and error rates are jointly considered. Moreover, the F statistics comparing these data across the two items are less than 1, meaning that the differences between means are inferior to the mean-square error of the data. For the same reasons, the Prism was chosen as the nonconjunction distractor to serve as a control for the Train top conjunction distractor.

This performance match between the conjunction and nonconjunction distractors to be used in Experiments 2–4 served to ensure that any surface-conjunction effect that might be observed in the latter experiments was not an artifact of a greater difficulty in discriminating the target from conjunction distractors than from nonconjunction distractors.

As an additional control of target/distractor similarity, we conducted an objective assessment of the similarity of the image of the target stimulus with that of each distractor selected for the conjunction and nonconjunction conditions. This was performed by calculating the normalized cross-correlation between the relevant image pairs normalized on mean pixel intensity in a location-independent manner. Specifically, the cross-correlations between image pairs were calculated for each possible location overlap between the two and selecting the highest correlation as reflecting their similarity; the higher the correlation, the greater the similarity. For the conjunction distractors, the similarity of the Asymmetrical cylinder with the target was 0.41, whereas that of the Train top was 0.78, for a mean similarity index of 0.59. The corresponding measures for the nonconjunction distractors were 0.45 for the Chamfer box and 0.75 for the Prism, giving a mean similarity index of 0.60. For all means and purposes, then, the target/distractor image similarities as assessed by this method are identical for the conjunction and nonconjunction conditions. The similarity of distractors was also precisely matched across the conjunction (0.39) and nonconjunction (0.38) distractors.

As described later, additional measures were used in Experiments 2–4 to further confirm the proper control of target/distractor similarity across conditions.

Experiment 2

The goal of Experiment 2 was to assess the surface-conjunction effect when the surfaces shared between the target and the conjunction distractors are so with respect to the contours of their 2-D image but not so in terms of their 3-D instantiation. Specifically, in the conjunction condition the target was made of a conjunction of surfaces that belong to the distractors. In contrast, in the nonconjunction condition the target had unique surfaces that were not shared with the distractors. The format of stimuli was as illustrated in Figure 1—that is, the objects had a textured surface and were realistically rendered to comprise shading information in order to offer significant and valid depth information.

Experiment 2 was divided in two separate tasks which were administered to distinct groups of participants. Experiment 2a served as the experimental task, where the number of instances of each distractor object for a particular condition (e.g., the Asymmetrical cylinder and the Train Top, for the conjunction condition) was as equal as possible on every trial. This means that in the conjunction condition of Experiment 2a, the target was truly made of a conjunction of distractor surfaces. In Experiment 2b, which served as a control task, participants were exposed on every trial to just one of the possible distractor objects, in as many instances as required to achieve the number of items to be displayed on that particular trial. On half the trials, one of the distractor objects for the condition was used; the other object was chosen for the other half of trials. Thus, in the conjunction condition of Experiment 2b, the same distractor objects were used as in Experiment 2a, but the target was never a conjunction of the surfaces of the distractors displayed along with it. It must be pointed out that implementing both Experiments 2a and 2b introduces a new factor in the experimental design (the same applies to Experiments 3 and 4). This factor is that of distractor heterogeneity, which has been demonstrated to negatively impact visual-search rates (Duncan & Humphreys, 1989, 1992). Thus, one should expect slower search rates with the heterogeneous distractors of Experiment 2a than with the homogenous distractors of Experiment 2b—a prediction that is fulfilled in Experiments 2–4. It is important to underline, however, that the manipulation of distractor heterogeneity is applied in precisely the same way for the conjunction and nonconjunction conditions.

Under the experimental design described, the conclusion that an actual surface-conjunction effect occurs will be based not only on the evidence that search performance is worse in the conjunction than in the nonconjunction condition (either in terms of greater RTs or error rates or a greater slope of RTs or error rates as a function of the number of items displayed) but also on the fact that this difference is unique to, or greater in, the experimental task (e.g., Experiment 2a) relative to the control task (e.g., Experiment 2b). From this design, then, the conclusion of a surface-conjunction effect will be protected from any artifact resulting from a mismatch of target/distractor similarity between the conjunction and nonconjunction conditions which may possibly remain despite the procedure used in Experiment 1 to select the nonconjunction distractors.

Given the preceding, the demonstration of a surface-conjunction effect in Experiment 2 would support the hypothesis of a surface-based code for shape representation and would suggest neglect (at least to some degree) of depth information in this surface code.

Method

Participants

Twenty-four volunteers (nine men and 15 women) aged between 19 and 39 years took part in the experiment. All were unaware of the purpose of the experiment, were neurologically intact, and had normal or corrected-to-normal visual acuity. No particular selection criteria were applied with respect to gender, manual dominance, or level of education. Participants were divided into two equal groups of 12 each. One group participated in Experiment 2a and the other in Experiment 2b.

Stimuli

The objects used in Experiment 2 are those marked Target, Asymmetrical cylinder, Train top, Chamfer box, and Prism in Figure 1. The cone serving as the target shares the contours of one of its surfaces (but not its texture gradients or shading) with each of the conjunction distractors—that is, the Asymmetrical cylinder and Train top. The other objects served as nonconjunction distractors and did not share any surface with the target. These distractors were chosen in Experiment 1 in order to match target/distractor similarity across the conjunction and nonconjunction distractors. The spatial extent of the stimuli varied between 1.33° and 2.5° of visual angle on the horizontal axis and between 0.95° and 1.71° on the vertical axis, from the viewing distance of 95 cm.

Procedure

Participants were instructed to indicate, on every trial, whether the target (i.e., the cone in Figure 1) was present in a display made of a variable number of items. The distractor sets were made of the conjunction or nonconjunction distractors, and the display size was of 3, 5, 7, or 9 items. Each stimulus on a trial was displayed randomly at one of 12 locations along an imaginary circle of 9.5° diameter, centered on fixation. On target-present trials (half of the trials), a single target was presented, whereas no target was displayed on target-absent trials. In Experiment 2a, the number of instances of each distractor object for the condition tested was made as equal as possible on every trial. In contrast, in Experiment 2b one of the two distractor objects for the condition tested was selected randomly and was replicated in the display as many times as necessary to achieve the required number of items for that trial. Whether in Experiment 2a or 2b, the total number of occurrences of each distractor object was equated within each block of trials (see later).

Trials began with a 500-ms fixation cross presented at the center of the computer screen. This was immediately followed by the stimulus display, which remained visible until the participant responded. The intertrial interval was 500 ms. Participants were instructed to respond as rapidly and as accurately as possible by pressing the appropriate response button of a response box interfaced with the computer controlling the experiment using the left or right index finger for the left and right response buttons, respectively. The side to which the target-present and target-absent buttons were assigned was counterbalanced across participants for both Experiments 2a and 2b. Participants were seated and had their head position restrained by a chin rest.

The experimental design involved the within-subject factors of distractors (2 levels: conjunction vs. non-

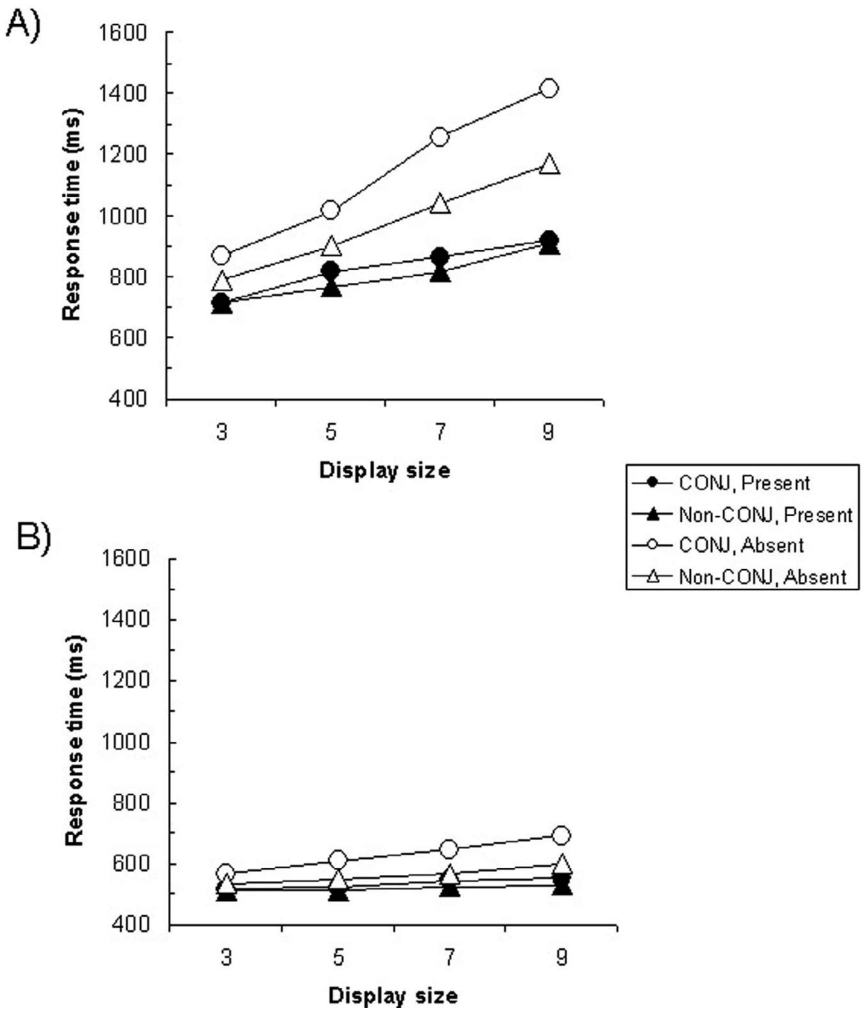

Figure 2. Average correct response times in Experiment 2 as a function of display size. (A) Experimental task (Experiment 2a); (B) control task (Experiment 2b).

conjunction), target presence (2 levels: present vs. absent), and display size (4 levels: 3, 5, 7, or 9 items), as well as the between-subjects factor of task (Experimental vs. Control; i.e., Experiment 2a vs. Experiment 2b). Each of the 16 possible within-subject conditions was presented in random order in two blocks of 384 trials (24 trials per condition per block), for a grand total of 768 trials per participant. These blocks were preceded by a sequence of 36 practice trials. For each participant, the experiment was run in a single session lasting approximately 40 min.

Results

Individual correct RTs that were more than 2.5 standard deviations from the participant's mean for that condition were excluded from data analysis. This resulted in the elimination of 463 trials (2.5% of all trials). The correlation between correct RTs and error rates was –0.13 (not significant), thus showing no speed/accuracy trade-off. Average correct RTs and error rates for each condition are illustrated in Figures 2 and 3, respectively.

Results summary

The main findings from the detailed data analyses described later are as follows. Correct RTs increased linearly with display size in all conditions, and they were larger with the conjunction than the nonconjunction distractors. On target-present trials, the latter effect was of the same magnitude in the experimental (Experiment 2a) and control (Experiment 2b) tasks. In contrast, on target-absent trials, correct RTs were greater, the effect of display size was also greater in the conjunction than the nonconjunction condition, and these differences were larger in the experimental (Experiment 2a) than in the control (Experiment 2b) task. Error rates were also larger in the conjunction than the nonconjunction condition, but this difference

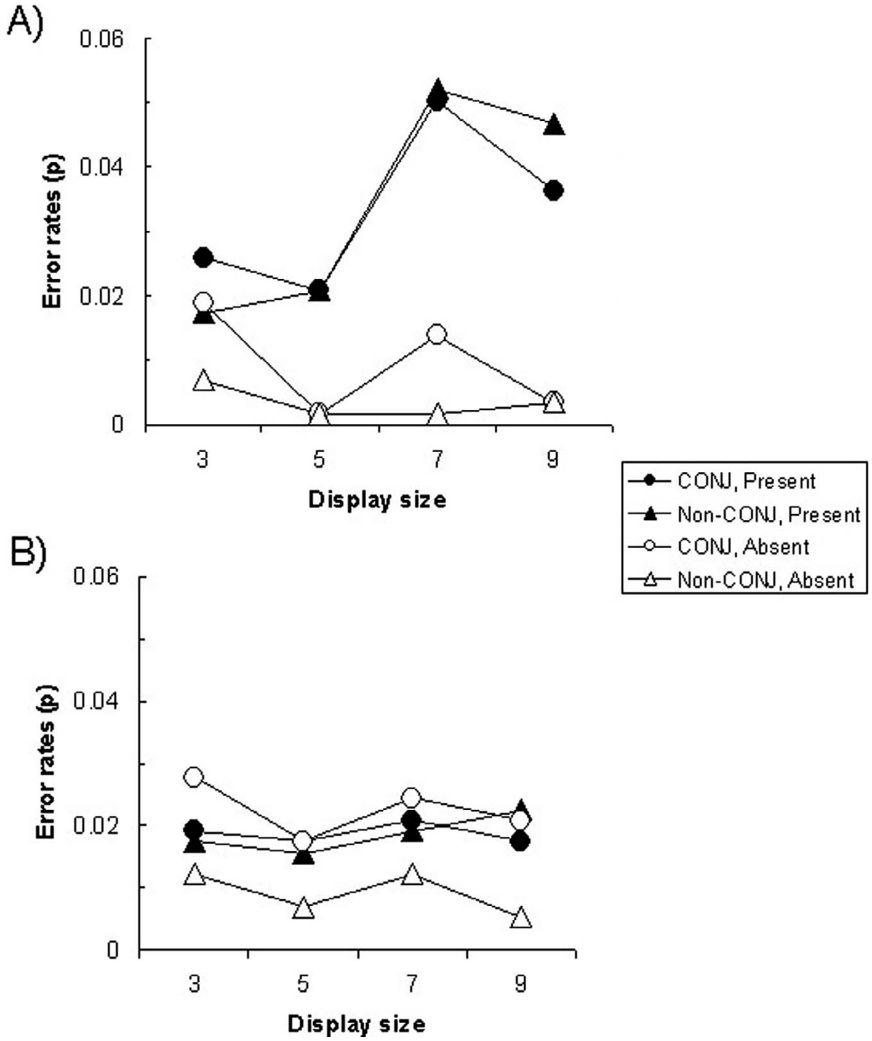

Figure 3. Average error rates in Experiment 2 as a function of display size. (A) Experimental task (Experiment 2a); (B) control task (Experiment 2b).

was of the same magnitude in the experimental (Experiment 2a) and control (Experiment 2b) tasks.

Correct RTs—Linear regressions

Linear-regression analyses of correct RTs as a function of display size were carried out to characterize the visual-search functions in each condition (Table 2). All regression functions are strictly linear. The ratio of target-present over target-absent regression slopes is 0.5 in the nonconjunction condition of Exp. 2a, which is congruent with a serial self-terminating search (but see Townsend, 1972, on the difficulty of discriminating between serial versus parallel search). In the remaining

	Target present			Target absent			
Condition	Intercept (ms)	Slope (ms/item)	R^2	Intercept (ms)	Slope (ms/item)	R^2	Present:absent slope ratio
Experimental task (Experiment 2a)							
Conjunction	630.3	33.0	0.97	571.1	94.6	0.99	0.35
Nonconjunction	612.9	31.7	0.97	591.8	63.7	1.00	0.50
Control task (Experiment 2b)							
Conjunction	494.4	6.5	0.96	504.4	20.5	1.00	0.32
Nonconjunction	500.6	3.2	0.89	496.6	10.9	0.95	0.29

Table 2. Linear regressions of correct response times as a function of display size in the conjunction and nonconjunction conditions of Experiment 2.

conditions, the ratios are weaker (from 0.29 to 0.35), implying that search rates on target-present trials are somewhat faster than what would be expected from target-absent search rates assuming a serial self-terminating search. In the experimental task (Experiment 2a), the slopes of correct RTs as a function of display size are substantially larger in the conjunction than in the nonconjunction condition on target-absent trials, but no difference is evident on target-present trials. In the control task (Experiment 2b), the slope differences between the conjunction and nonconjunction conditions are weak but tend to favor the latter.

Correct RTs—ANOVA

A four-way analysis of variance (ANOVA) with the within-subject factors of distractors (2 levels: conjunction vs. nonconjunction), target presence (2 levels: present vs. absent), and display size (4 levels: 3, 5, 7, or 9 items) and the between-subjects factor of task (Experiment 2a vs. 2b) was performed on correct RTs. This analysis showed several main effects and interactions, including a significant four-way Distractors × Target presence × Display size × Task interaction, $F(3, 66) = 3.9$, $p < 0.05$. In consequence of this interaction, the experimental design was divided according to target-present and target-absent trials to compare the joint effects of distractors and display size across tasks.

Target-present trials: For target-present trials, the analysis revealed main effects of distractors, $F(1, 22) = 9.4$, $p < 0.01$; display size, $F(3, 66) = 46.8$, $p < 0.001$; and task, $F(1, 22) = 409.1$, $p < 0.001$; as well as a significant Display size × Task interaction, $F(3, 66) = 25.9$, $p < 0.001$. These effects indicate longer RTs in the conjunction than in the nonconjunction condition, increasing RTs with the number of items displayed, and a greater effect of display size in Experiment 2a than 2b (see Figure 2 and Table 2). All other effects were not significant ($p > 0.05$). The observations for target-present trials fail to verify the conditions defined earlier for a surface-conjunction effect. Thus, while RTs are longer in the conjunction than the nonconjunction conditions, this difference is not significantly greater in the experimental than in the control task (note the lack of a Distractors × Task interaction).

Target-absent trials: The analysis of target-absent trials, however, does provide evidence for a surface-conjunction effect in Experiment 2. This analysis demonstrates significant main effects of distractors, $F(1, 22) = 30.6$, $p < 0.001$; display size, $F(3, 66) = 32.7$, $p < 0.001$; and task, $F(1, 22) = 219.7$, $p < 0.001$. In addition, a number of the two-way interactions were significant: Distractors × Display size, $F(3, 66) = 14.4$, $p < 0.001$; Distractors × Task, $F(1, 22) = 5.8$, $p < 0.05$; and Display size × Task, $F(3, 66) = 14.8$, $p < 0.001$. These indicate a greater effect of display size in the conjunction than the nonconjunction condition, a greater performance difference between the conjunction and nonconjunction conditions in Experiment 2a than in 2b, and a greater effect of display size in Experiment 2a than in 2b (see Table 2 and Figure 2). The three-way Distractors × Display size × Task interaction was also significant, $F(3, 66) = 4.3$, $p < 0.01$. A breakdown of that interaction indicates significant Distractors × Display size interactions for both Experiment 2a, $F(3, 33) = 11.7$, $p < 0.001$, and Experiment 2b, $F(3, 33) = 2.9$, $p < 0.05$. However, the interaction is substantially greater for Experiment 2a than 2b, thereby accounting for the three-way Distractors × Display size × Task interaction on target-absent trials. In summary, the RT evidence on target-absent trials meets the criteria for a surface-conjunction effect. Thus, the cost associated with the conjunction condition (relative to the nonconjunction condition) on overall RTs (165 ms in Experiment 2a, 65 ms in Experiment 2b), as well as on the slope of RTs as a function of number of items (31 ms/item in Experiment 2a, 10 ms/item in Experiment 2b), is greater in Experiment 2a (experimental task) than in Experiment 2b (control task).

Error rates—ANOVA

Error rates were analyzed using an ANOVA with the same design as for RTs. This analysis revealed several main effects and two-way interactions involving the factors of target presence, display size, and task, which were qualified by the significant triple Target presence × Display size × Task interaction, $F(3, 66) = 3.6$, $p < 0.05$. This interaction rests on the fact that for the experimental task (Experiment 2a), error rates were greater on target-present than on target-absent trials, $F(1, 11) = 47.4$, $p < 0.001$, and they varied in an irregular manner as a function of display size (see Figure 3). In contrast, in the control task (Experiment 2b), there was no effect of target presence, $F(1, 11) = 1.1$ (not significant), or display size, $F(3, 33) < 1$, or an interaction between them, $F(3, 33) < 1$. The general ANOVA applied on error rates also showed a main effect of distractors as well as a Distractors × Target presence interaction. A breakdown of the interaction revealed significantly larger error rates in the conjunction than in the nonconjunction condition on target-absent trials, $F(1, 23) = 11.5$, $p < 0.005$, but no effect on target-present trials, $F(1, 23) < 1$. The analysis of error rates revealed no difference between tasks on the magnitude of the effect of distractors, $F(1, 22) = 1.1$ (not significant), nor on how it modulates the effect of display size, $F(3, 66) < 1$.

Discussion

The main finding from Experiment 2 is the evidence for a surface-conjunction effect on RTs on target-absent trials. Indeed, RTs were longer, and their slope as a function of display size larger, in the conjunction than in the nonconjunction condition, and this difference was greater in the experimental task (Experiment 2a) than in the control task (Experiment 2b). In simpler terms, this means that visual search was slower when the target was a conjunction of surfaces constituting the distractors displayed along with it than when it was not. This performance cost is not simply an artifact of target/distractor similarity, which was fully controlled. As argued earlier, this observation supports the hypothesis that internal representations coding visual shapes as collections of surfaces were involved in performing the visual-search task.

It should be noted that the surface-conjunction effect occurred in Experiment 2 even though the depth cues offered by texture and shading indicated that the orientation in depth of the surfaces making up the conjunction target was different from the corresponding surfaces in the distractors. This evidence suggests that the surface-based code for shape representation ignores depth information—that is, it takes into account only 2-D contour information. Experiment 3 assessed this hypothesis further.

Experiment 3

The results of Experiment 2 were obtained using stimuli that were relatively rich in terms of depth information. Nevertheless, the depth information available in the stimuli is contradicted by the null binocular disparity, which signals that the stimuli displayed are flat. It could thus be argued that the surface-based code suggested by Experiment 2 does take into account depth information, but that this information must be congruent across depth cues to register. Alternatively, it could be argued that the surface-based code obeys specifically binocular disparity information in order to represent the depth orientation of surfaces. Experiment 3 assessed these possibilities by replicating Experiment 2, but using anaglyphic stimuli that contain valid binocular disparity depth information with respect to the objects depicted.

Method

Participants

Twenty-four volunteers (eight men and 16 women) aged between 19 and 42 years took part in the experiment. All were unaware of the purpose of the experiment, were neurologically intact, and had normal or corrected-to-normal visual acuity. No particular selection criteria were applied with respect to gender, manual dominance, or level of education. Participants were divided into two equal groups of 12 participants each, one taking part in Experiment 3a and the other Experiment 3b.

Stimuli

Experiment 3 used the same set of objects as Experiment 2, which were presented as anaglyphs in order to offer stereoscopic depth information. Two renderings of each object were produced by rotating the stimulus views depicted in Figure 1 by 1.75° to the left or right around a vertical axis passing through the center of the object. This angle of rotation simulates the views of the object for the right and left eye, respectively, assuming an interocular distance of 5.8 cm and a viewing distance of 95 cm (i.e., that used in the present experiment). The stimulus views were then fused to produce red/cyan anaglyphs which were the actual stimuli used in Experiment 3. The stimuli were viewed using red/cyan glasses (Rainbow Symphony, Reseda, CA) to give observers the impression of stereoscopic depth. Stimuli were displayed over a neutral gray background of medium intensity (instead of the white background used in Experiment 2) in order to avoid phantom images from portions of the stimulus that were designed to be viewed only by the contralateral eye.

Procedure

Apart from the fact that participants were required to wear the red/cyan glasses throughout the duration of the experiment, the experimental design and procedures were identical to those of Experiment 2.

Results

Individual correct RTs that were more than 2.5 standard deviations from the participant's mean for that condition were excluded from data analysis. This resulted in the elimination of 495 trials (2.7% of all trials). The correlation between correct RTs and error rates was −0.03 (not significant), thus indicating no speed/accuracy trade-off. Average correct RTs and error rates for each condition are illustrated in Figures 4 and 5, respectively.

Results summary

The results of Experiment 3 show that correct RTs increased linearly with display size in all but one

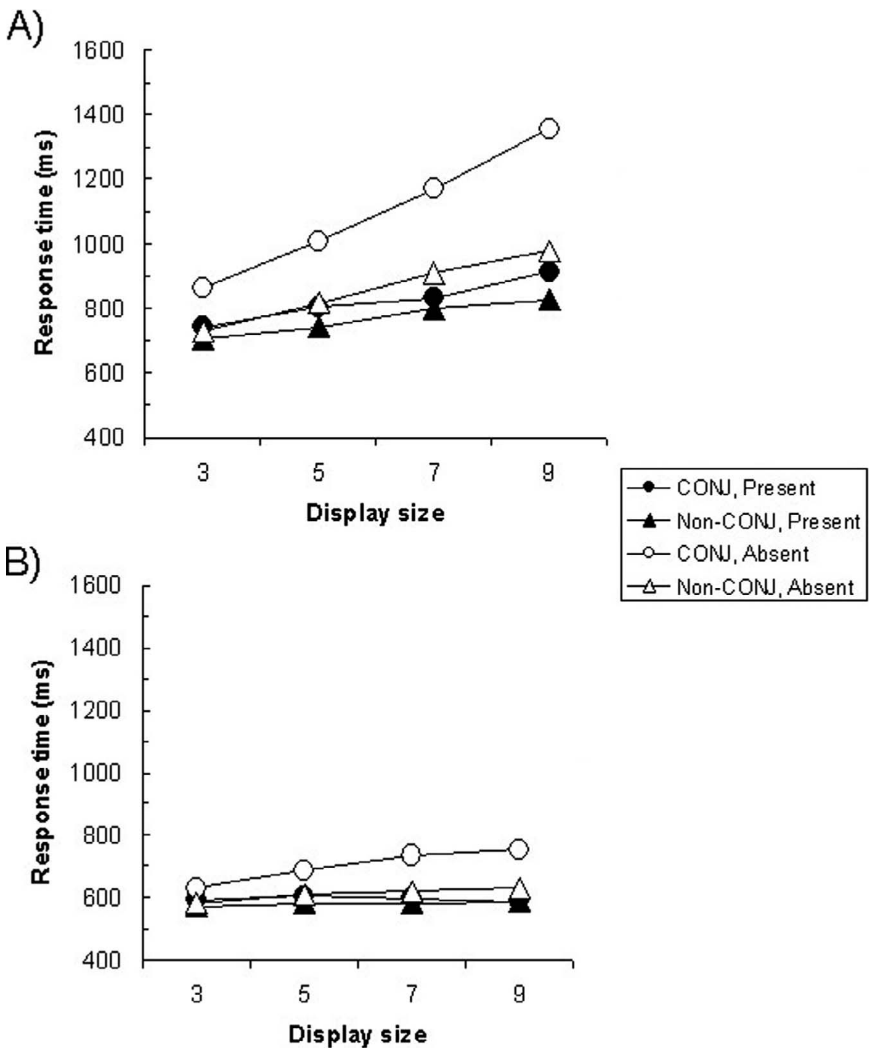

Figure 4. Average correct response times in Experiment 3 as a function of display size. (A) Experimental task (Experiment 3a); (B) control task (Experiment 3b).

condition, for which the slope was very weak. On both target-present and target-absent trials, correct RTs and the effect of display size on correct RTs were greater in the conjunction than the nonconjunction condition, and these effects had a greater magnitude in the experimental (Experiment 3a) than in the control (Experiment 3b) task. No conjunction effect occurred on error rates. These findings are demonstrated by the data analyses described in the following.

Correct RTs—Linear regressions

Linear-regression analyses of correct RTs as a function of display size were carried out to characterize the visual-search functions in each condition (Table 3). All regression functions are very linear except for target-present trials in the conjunction condition of the control task (Experiment 3b). The R^2 statistic is particularly low in this case, but it is associated with an effect of display size that is practically null. The ratio of target-present over target-absent regression slopes is close to 0.5 in the nonconjunction condition of the experimental task (Experiment 3a), congruent with a serial self-terminating search (but see Townsend, 1972). In the other conditions, however, the ratio is weaker (0.34). In the conjunction condition of the experimental task and the nonconjunction condition of the control task, this simply means that the search rates on target-present trials are somewhat faster than what would be expected on the basis of target-absent search rates if one were to assume a serial self-terminating search. In the conjunction condition of the control task, the near-zero ratio is explained by the null slope on target-present trials. In the experimental task (Experiment 3a), the slopes of correct RTs as a function of display size are substantially larger in the conjunction than in the

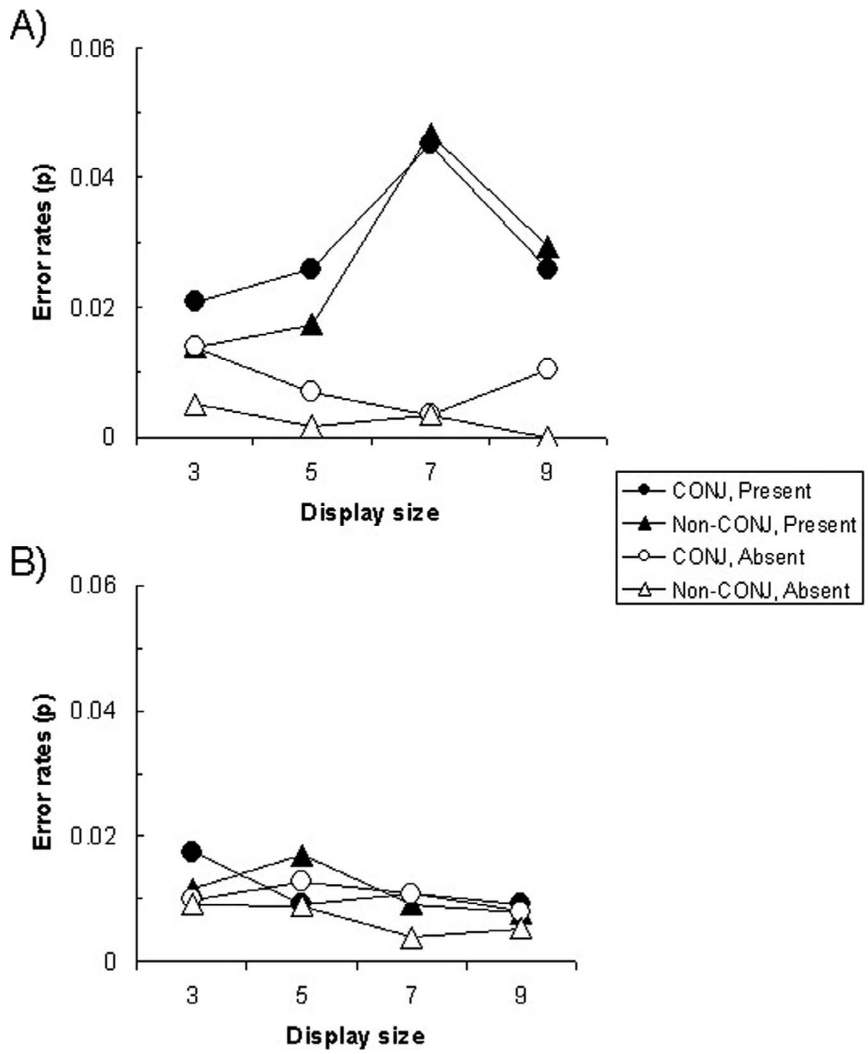

Figure 5. Average error rates in Experiment 3 as a function of display size. (A) Experimental task (Experiment 3a); (B) control task (Experiment 3b).

nonconjunction condition on target-absent trials, but the difference is weaker on target-present trials. In the control task (Experiment 3b), the slope differences between the conjunction and nonconjunction conditions are weaker than in the experimental task, and on target-present trials the difference tends to favor the conjunction condition.

Correct RTs—ANOVA

A four-way ANOVA with the within-subject factors of distractors (2 levels: conjunction vs. nonconjunction), target presence (2 levels: present vs. absent), and display size (4 levels: 3, 5, 7, or 9 items) and the between-subjects factor of task (Experiment 3a vs. 3b)

	Target present			Target absent			
Condition	Intercept (ms)	Slope (ms/item)	R^2	Intercept (ms)	Slope (ms/item)	R^2	Present:absent slope ratio
Experimental task (Experiment 3a)							
Conjunction	656.3	27.6	0.97	607.1	82.0	1.00	0.34
Nonconjunction	642.5	20.9	0.98	601.3	42.8	1.00	0.49
Control task (Experiment 3b)							
Conjunction	599.9	−0.8	0.08	578.6	20.5	0.95	−0.04
Nonconjunction	566.0	2.6	0.88	562.7	8.0	0.92	0.32

Table 3. Linear regressions of correct response times as a function of display size in the conjunction and nonconjunction conditions of Experiment 3.

was performed on correct RTs. This analysis showed several main effects and interactions, including several significant three-way interactions: Distractors × Display size × Task, $F(3, 66) = 12.5$, $p < 0.001$; Target presence × Display size × Task, $F(3, 66) = 9.5$, $p < 0.001$; and Distractors × Target presence × Display size, $F(3, 66) = 15.6$, $p < 0.001$. In order to obtain a properly detailed picture of the effects on correct RTs, the overall experimental design was broken down into analyses of the joint effects of distractors and display size across tasks separately for target-present and target-absent trials.

Target-present trials: On target-present trials, all main effects and interactions were significant ($p < .05$) except for the Distractors × Display size interaction, $F(3, 66) = 1.4$ (not significant). The significant Distractors × Task interaction, $F(1, 22) = 7.8$, $p < 0.05$, indicates that the RT cost associated with the conjunction condition was greater in the experimental task (54 ms; Experiment 3a) than in the control task (13 ms; Experiment 3b). To follow up on the significant three-way Distractors × Display size × Task interaction, $F(3, 66) = 4.2$, $p < 0.01$, the joint effects of distractors and display size were analyzed separately for each task. For the experimental task (Experiment 3a), the Distractors × Display size interaction was significant, $F(3, 33) = 3.0$, $p < 0.05$, indicating a significantly greater effect of display size in the conjunction than the nonconjunction condition (see Table 3 and Figure 4). For the control task (Experiment 3b), the Distractors × Display size interaction was also significant, $F(3, 33) = 3.5$, $p < 0.05$. In this case however, the slope of RTs as a function of display size is weaker in the conjunction than in the nonconjunction condition (see Table 3).

Target-absent trials: On target-absent trials, the analysis of RTs according to the Distractors × Display size × Task interaction showed that all main effects and interactions were significant ($p < 0.05$). The significant Distractors × Task interaction, $F(1, 22) = 18.7$, $p < 0.001$, indicates that the RT cost associated with the conjunction condition was greater in the experimental task (241 ms; Experiment 3a) than in the control task (91 ms; Experiment 3b). The joint effects of distractors and display size were analyzed separately according to task to break down the significant three-way Distractors × Display size × Task interaction, $F(3, 66) = 8.0$, $p < 0.001$. The Distractors × Display size interaction was significant for both the experimental task, $F(3, 33) = 30.2$, $p < 0.001$, and the control task, $F(3, 33) = 3.5$, $p < 0.05$. However, as is apparent in Table 3, the magnitude of the conjunction cost on the effect of display size was greater in the experimental task (slope difference of 39.2 ms/item) than in the control task (slope difference of 12.5 ms/item), which accounts for the significant Distractors × Display size × Task interaction on target-absent trials.

In summary, the RT results for both target-present and target-absent trials met the criteria for a surface-conjunction effect. Thus, the conjunction condition is associated with worse performance than the non-conjunction condition (i.e., greater overall RTs and a greater increase of RTs as a function of display size) and this cost is greater in, or unique to, the experimental task (relative to the control task).

Error rates—ANOVA

Error rates were analyzed using an ANOVA with the same design as for RTs. The main effects of target presence, $F(1, 22) = 31.8$, $p < 0.001$, was significant (greater error rates on target-present than target-absent trials), along with multiple two-way interactions: Target presence × Task, $F(1, 22) = 18.7$, $p < 0.001$; Target presence × Display size, $F(3, 66) = 4.8$, $p < 0.005$; and Display size × Task, $F(3, 66) = 4.7$, $p < 0.01$. These interactions were qualified by the significant three-way Target presence × Display size × Task interaction, $F(3, 66) = 5.9$, $p < 0.005$. Simple effects of display size and task were carried out separately on target-present and target-absent trials. On target-present trials, the Display size × Task interaction was significant, $F(3, 66) = 6.7$, $p < 0.005$, which indicates greater error rates in the experimental (Experiment 3a) than in the control task (Experiment 3b). In contrast, the analysis of the joint effects of display size and task on target-absent trials revealed no significant effect or interaction ($p > 0.05$). In summary, the analysis of error rates indicated only that participants made more errors on target-present than target-absent trials in the experimental task (Experiment 3a), whereas no such difference occurred in the control task (Experiment 3b).

Discussion

The results of Experiment 3 replicate the surface-conjunction effect of Experiment 2. Specifically, RTs were longer and the slope of RTs as a function of display size was larger in the conjunction than in the nonconjunction condition, and this difference was greater in the experimental task (Experiment 3a) than in the control task (Experiment 3b).

Such observations indicate that the surface code responsible for the conjunction effect investigated here fails to take into account the fact that the 2-D regions shared between the target and distractors are completely different in their 3-D instantiations. This shows that a surface-based representation of visual shape does not integrate depth information and is based instead on 2-Dcontour information.

Experiment 4

The results of Experiments 2 and 3 converge to show that a surface-based code for shape representation does not contain depth information. Indeed, had this been so, one would expect no significant surface-conjunction effect in these experiments. Still, under the assumption that the surface-based code might be capable of processing depth information to some degree, albeit imperfectly, it could be argued that the surface-conjunction effect in Experiments 2 and 3 was weaker than if the depth information presented had been much poorer.

Experiment 4 was designed to assess this possibility by using stimuli with very impoverished depth information. Thus, stimuli were displayed without binocular disparity information and the objects were rendered without any surface texture, which instead had a matte gray surface. With such stimuli, the only reliable depth information available is that of shading. In other respects, the methods applied in Experiment 4 are identical to those of Experiments 2 and 3.

Method

Participants

Twenty-four volunteers (nine men and 15 women) aged between 18 and 42 years took part in the experiment. All were unaware of the purpose of the experiment, were neurologically intact, and had normal or corrected-to-normal visual acuity. No particular selection was applied with respect to gender, manual dominance, or level of education. Participants were divided into two equal groups of 12 participants each, one taking part in Experiment 4a and the other Experiment 4b.

Stimuli

Experiment 4 used the same set of objects as Experiment 2, which were rendered with matte gray surfaces devoid of any texture. Thus, the only depth information available in these stimuli was provided by shading. Stimuli were displayed over a white background.

As for the textured stimuli used in the preceding experiments, an objective assessment of the similarity between the image of the target and of those of the distractors was calculated using the normalized cross-correlation between the relevant image pairs normalized on mean pixel intensity. For the conjunction distractors, the similarity of the Asymmetrical cylinder with the target was 0.31, whereas that of the Train top was 0.65, for a mean similarity index of 0.48. The corresponding measures for the nonconjunction distractors were 0.36 for the Chamfer box and 0.57 for the Prism, giving a mean similarity index of 0.46. Thus, the similarity of the distractors to the target was precisely matched between the conjunction and nonconjunction conditions. The similarity between distractors in the conjunction (0.31) and nonconjunction (0.35) conditions was also very close.

Procedure

The experimental design and procedures were identical to those of Experiment 2.

Results

Individual correct RTs that were more than 2.5 standard deviations from the participant's mean for that condition were excluded from data analysis. This resulted in the elimination of 466 trials (2.6% of all trials). The correlation between correct RTs and error rates was −0.12 (not significant), thus indicating no significant speed/accuracy trade-off. Average correct RTs and error rates in each condition of Experiment 4 are shown in Figures 6 and 7, respectively.

Results summary

Except for one condition where the slope was very weak, correct RTs increased linearly with display size. The main outcomes of the analyses of variance applied on correct RTs and error rates are as follows. On both target-present and target-absent trials, correct RTs as well as the effect of display size on correct RTs were greater in the conjunction than the nonconjunction condition, and these effects had a greater magnitude in the experimental (Experiment 4a) than in the control (Experiment 4b) task. On target-present trials, error rates increased faster with display size in the conjunction than in the nonconjunction condition. On target-absent trials, error rates were greater in the conjunction than in the nonconjunction condition. However, the magnitude of these conjunction effects on error rates did not differ between the experimental (Experiment 4a) and control (Experiment 4b) tasks. Additional analyses comparing outcomes across experiments showed that the pattern of results pertaining to the demonstration of a surface-conjunction effect did not differ significantly between Experiments 2, 3, and 4.

Correct RTs—Linear regressions

Linear-regression analyses of correct RTs as a function of display size were carried out to characterize the visual-search functions in each condition (Table 4). Except for the nonconjunction condition of the control

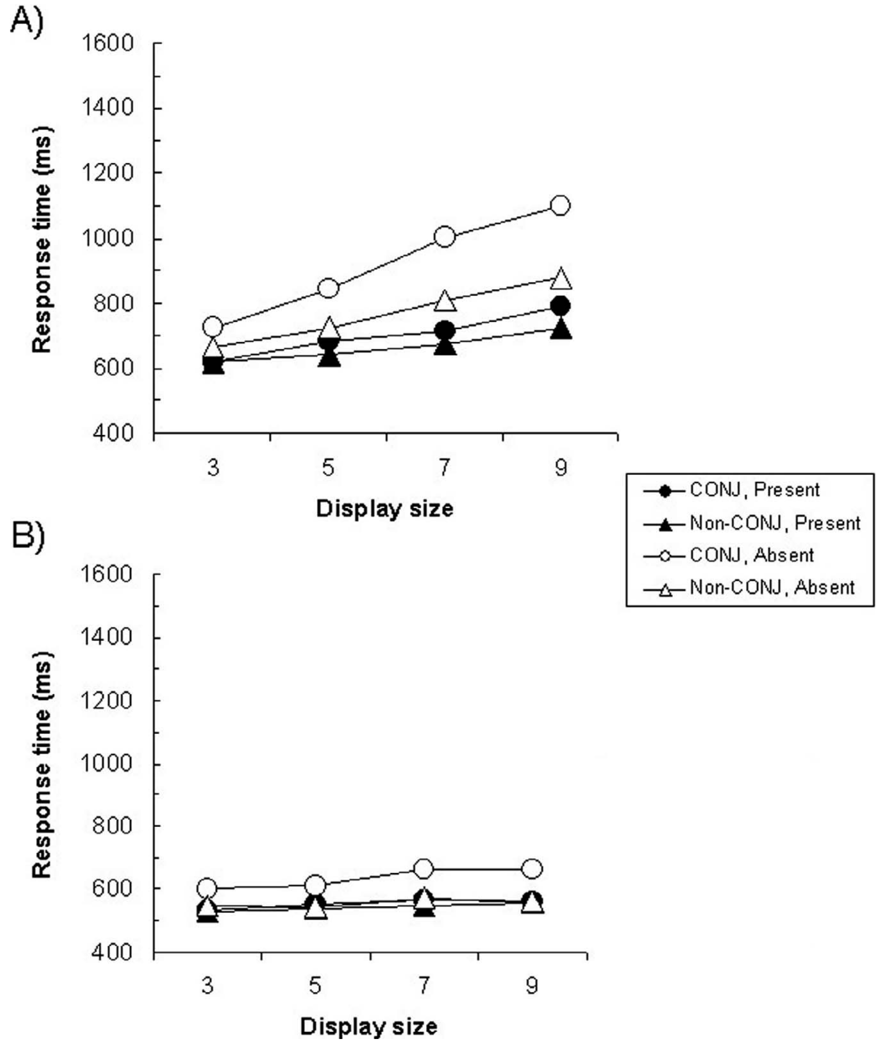

Figure 6. Average correct response times in Experiment 4 as a function of display size. (A) Experimental task (Experiment 4a); (B) control task (Experiment 4b).

task (Experiment 4b), all regression functions are linear and the ratios of target-present regression slopes over target-absent are about 0.5. This is congruent with a serial self-terminating search (but see Townsend, 1972). For the nonconjunction condition of the control task (Experiment 4b), the regression function on target-absent trials is relatively nonlinear and the slope is weaker than on target-present trials, hence the particular present/absent slope ratio. In the experimental task (Experiment 4a), the slopes of RTs as a function of display size are notably larger in the conjunction than in the nonconjunction condition for both target-present and target-absent trials. The corresponding contrasts for the control task (Experiment 4b) are much weaker.

Correct RTs—ANOVA

Similarly to Experiments 2 and 3, correct RTs were analyzed using a four-way ANOVA with the within-subject factors of distractors (2 levels: conjunction vs. nonconjunction), target presence (2 levels: present vs. absent), and display size (4 levels: 3, 5, 7, or 9 items) and the between-subjects factor of task (Experiment 4a vs. 4b). All main effects and interactions were significant (all *p*s < 0.05), except for the Distractors $\times$ Target presence $\times$ Task interaction, $F(1, 22) = 3.7$ (not significant), and the four-way interaction, $F(3, 66) = 1.5$ (not significant). As in the previous experiments, more detailed analyses were conducted to assess the joint effects of distractors and display size across tasks, separately for target-present and target-absent trials. *Target-present trials*: On target-present trials, all main effects and interactions were significant, including the two-way Distractors $\times$ Task interaction, $F(1, 22) = 7.5$, $p < 0.05$, and the three-way Distractors $\times$ Display size $\times$ Task interaction, $F(3, 66) = 3.6$, $p < 0.05$. Simple effects of distractors and display size were therefore analyzed separately for each task. For the experimental task (Experiment 4a), there were significant main effects of distractors, $F(1, 11) = 15.9$, $p < 0.005$, and

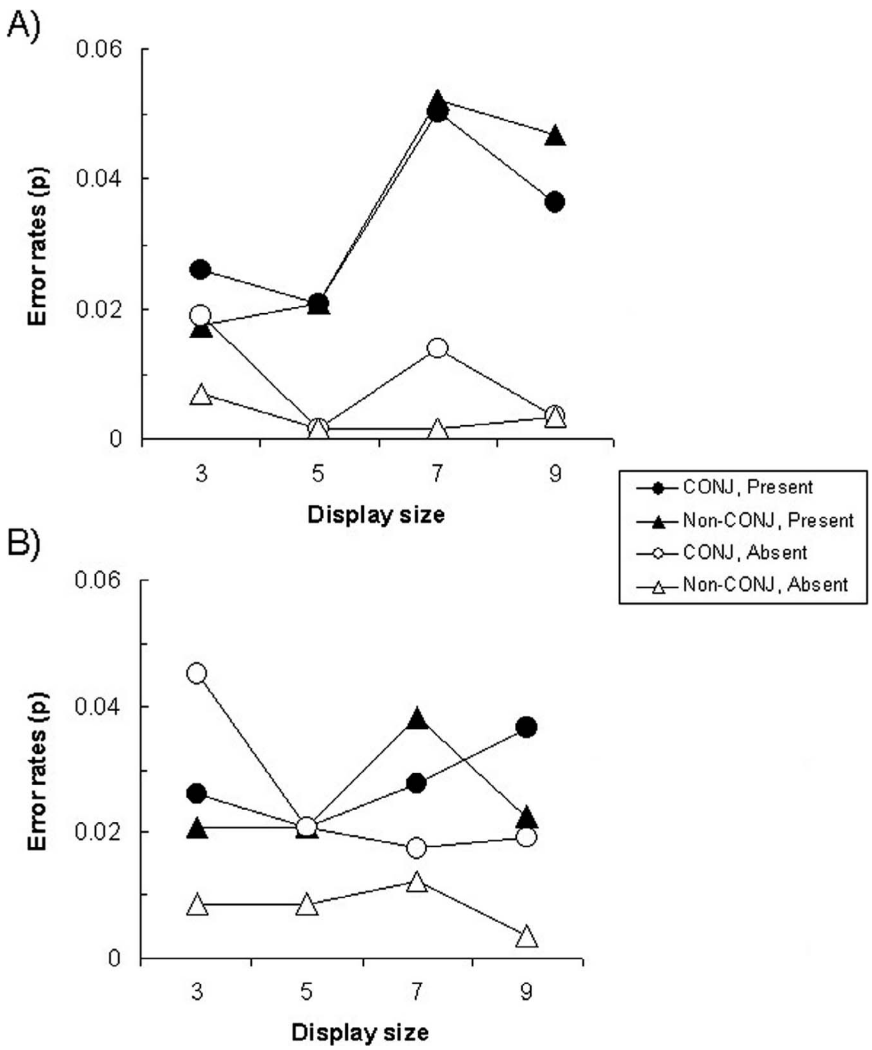

Figure 7. Average error rates in Experiment 4 as a function of display size. (A) Experimental task (Experiment 4a); (B) control task (Experiment 4b).

display size, $F(3, 33) = 37.7$, $p < 0.001$, along with the Distractors × Display size interaction, $F(3, 33) = 6.6$, $p < 0.005$. This interaction reflects the fact that the increase of RTs as a function of display size was greater in the conjunction than the nonconjunction condition, as can be seen in Table 4 and Figure 6A. For the control task (Experiment 4b), there were significant main effects of distractors, $F(1, 11) = 9.8$, $p < 0.05$, and display size, $F(3, 33) = 10.2$, $p < 0.001$, the Distractors × Display size interaction was not significant, $F(3, 33) < 1$. Thus, on target-present trials in the control task, correct RTs were slightly longer in the conjunction than in the nonconjunction condition, and they increased with display size. However, this effect of display size

	Target present			Target absent			
Condition	Intercept (ms)	Slope (ms/item)	R^2	Intercept (ms)	Slope (ms/item)	R^2	Present:absent slope ratio
Experimental task (Experiment 4a)							
Conjunction	534.3	27.7	0.98	532.7	64.1	0.99	0.43
Nonconjunction	555.8	18.0	0.97	552.4	36.1	0.99	0.50
Control task (Experiment 4b)							
Conjunction	521.3	5.4	0.83	560.7	12.2	0.84	0.44
Nonconjunction	515.5	4.7	0.99	537.2	3.4	0.53	1.36

Table 4. Linear regressions of correct response times as a function of display size in the conjunction and nonconjunction conditions of Experiment 4.

did not vary according to distractors (see Table 4 and Figure 6B). In summary, with respect to our main purpose of assessing the surface-conjunction effect, these analyses indicate that for target-present trials, both the effect of distractors (37 ms vs. 10 ms) and its impact on the effect of display size (9.7 ms/item vs. 0.7 ms/item) were greater in the experimental than the control task (Experiment 4a vs. 4b).

Target-absent trials: On target-absent trials, the pattern of effects on correct RTs is similar to that of target-present trials, but the magnitude of the surface-conjunction effect is greater. The simple effects of distractors, display size, and task on target-absent trials revealed all main effects and interactions to be significant, including the two-way Distractors × Task interaction, $F(1, 22) = 10.8$, $p < 0.005$, and the three-way Distractors × Display size × Task interaction, $F(3, 66) = 6.9$, $p < 0.005$. For the experimental task (Experiment 4a), there were significant main effects of distractors, $F(1, 11) = 113.2$, $p < 0.001$; display size, $F(3, 33) = 34.2$, $p < 0.001$; and the Distractors × Display size interaction, $F(3, 33) = 23.4$, $p < 0.001$. These indicate that RTs increased as a function of display size and that RTs as well as the effect of display size were greater on conjunction than nonconjunction trials (see Table 4 and Figure 6A). For the control task (Experiment 4b), there were significant main effects of distractors, $F(1, 11) = 19.7$, $p < 0.005$; display size, $F(3, 33) = 9.7$, $p < 0.001$; and the Distractors × Display size interaction, $F(3, 33) = 3.9$, $p < 0.05$. Again, these effects indicate that RTs increased as a function of display size and that RTs and the effect of display size were greater with conjunction than nonconjunction distractors (see Table 4 and Figure 6B). It may be noted however, that the effect of distractors (Experiment 4a: 148 ms; Experiment 4b: 76 ms) and its impact on the effect of display size (Experiment 4a: 28.0 ms/item; Experiment 4b: 8.7 ms/item) were greater in the experimental than in the control task.

Error rates—ANOVA

Error rates were analyzed using the same design as for RTs. There were significant main effects of distractors, $F(1, 22) = 13.4$, $p < 0.005$; target presence, $F(1, 22) = 25.3$, $p < 0.001$; display size, $F(3, 66) = 4.6$, $p < 0.01$; and task, $F(1, 22) = 55.6$, $p < 0.001$. These indicate greater error rates with conjunction than with nonconjunction distractors (3.1% vs. 1.9%), on target-present than on target-absent trials (3.4% vs. 1.6%), and in the experimental task than in the control task (2.8% vs. 2.2%), as well as increasing error rates as a function of display size, except for a dip with five-item displays (2.4%, 1.8%, 2.7%, and 2.9% with display sizes of 3, 5, 7, and 9, respectively). The two-way Target presence × Task interaction was significant, $F(1, 22) = 5.0$, $p < 0.001$, indicating that the effect of target presence was greater in the experimental than in the control task (effect size of 2.5% vs. 1.0%). The significant Display size × Task interaction, $F(3, 66) = 3.7$, $p < 0.05$, shows that the pattern of error rates as a function of display size varied according to task (roughly, increasing with display size for the experimental task and trending irregularly downward with increasing display size for the control task). There were significant two-way interactions of Distractors × Display size, $F(3, 66) = 3.5$, $p < 0.05$, and Target presence × Display size, $F(3, 66) = 11.3$, $p < 0.001$, but they were qualified by the three-way Distractors × Target presence × Display size interaction, $F(3, 66) = 3.2$, $p < 0.05$.

Target-present trials: Simple effects of this interaction showed, for target-present trials, a significant main effect of display size, $F(3, 66) = 8.6$, $p < 0.001$, and a Distractors × Display size interaction, $F(3, 66) = 4.2$, $p < 0.01$. The latter interaction indicates that the pattern of roughly increasing error rates with display size on target-present trials was of a greater magnitude with conjunction than nonconjunction distractors.

Target-absent trials: On target-absent trials, only the main effects of distractors, $F(1, 22) = 13.9$, $p < 0.005$, and display size, $F(3, 66) = 8.7$, $p < 0.001$, were significant. Error rates were greater with conjunction than nonconjunction distractors (2.4% vs. 0.8%), and they decreased slightly with increasing display size (see Figure 7).

Between-experiments contrasts

An important issue Experiment 4 was designed to address is whether a radical reduction of depth cues relative to those used in Experiment 3 leads to a magnification of the surface-conjunction effect. The results clearly falsify this hypothesis. Thus, an additional analysis of correct RTs was conducted to determine whether the magnitude of the surface-conjunction effect demonstrated in Experiment 3 and 4 differs across experiments. This analysis involved the within-subject factors of distractors, target presence, and display size, and the between-subjects factors of task and experiment. To demonstrate that the magnitude of the surface-conjunction effect differs across experiments, what would be required is a significant interaction that jointly involves the factors of distractors, task, and group. None of the interactions jointly involving these factors was significant (all $ps < 0.05$). A similar contrast between Experiments 2 and 4 was performed with the same outcome.

Discussion

The results of Experiment 4 provide an additional replication of the surface-conjunction effect demonstrated in Experiments 2 and 3. Thus, correct RTs and the effect of display size on RTs were greater in the conjunction than in the nonconjunction condition, and these effects had a greater magnitude in the experimental than in the control task (Experiments 4a vs 4b).

An additional finding from Experiment 4 is that a major reduction in the depth information available does not alter the magnitude of the performance cost associated with the conjunction condition compared to stimulus displays that were notably richer with respect to the available depth cues, such as we used in Experiments 2 and 3.

These observations indicate that the surface-based code mediating the surface-conjunction effect demonstrated here disregards depth information, and that the occurrence of this phenomenon is unaltered by the manipulations of the richness of depth cues that were performed here.

General discussion

We used the visual-search task to assess the validity of surfaces as elementary features for shape representation in humans and to determine the sensitivity of this code to depth information. In all experiments, we contrasted a condition where the visible surfaces of the target were a conjunction of distractor surfaces (conjunction condition) to another where target surfaces were unique in the display (nonconjunction condition). Crucially, given the triple-control measures applied here to match target/distractor similarities across conditions (see Discussion under Experiment 1 and Method under Experiments 2 and 4), the present assessment of a surface-conjunction effect is protected against any artifact pertaining to a potential mismatch across conditions in the discriminability of the target against individual distractors. This means that any performance difference across conditions must be attributable to whether the target is made of a conjunction of distractor surfaces or not.

In all experiments assessing the surface-conjunction effect (Experiments 2–4), visual-search performance was poorer when the target was a conjunction of distractor surfaces than when it was not. This poorer performance was manifest in terms of slower search rates and greater RTs, which we demonstrated in all experiments. The analysis of error rates failed to show any relevant evidence for or against the hypothesis of a surface-based code for shape representation. Overall then, the present findings support the notion of a surface-based code for shape representation. Thus, under this code, a conjunction target in our experiments had no unique property in the display and its reliable detection required the integration of two or more surfaces, thereby leading to a performance cost relative to the nonconjunction condition. Something that may be noted regarding the present demonstration is that it is based on a limited number of stimuli. Future research interested in the issue should aim at multiplying the number of different stimulus instances that are used to assess the application of a surface-based code to represent shape in human vision. This concern however, is mitigated by the fact that a number of other studies using different tasks and a large variety of stimuli have also reported potent evidence in support of a surface code (Leek et al., 2005; Leek et al., 2008; Leek et al., 2018; Reppa et al., 2015; Reppa & Leek, 2019).

In addition to providing evidence for a surface-based code for shape representation, the present study also contributes to specifying the properties of this code. Specifically, the present data indicate that the surface code appears to be blind to depth and to exclusively represent shape in terms of 2-D information. Thus, Experiments 2–4 show that the surface-conjunction effect occurs even though the surfaces shared between the target and distractors in the 2-D image are not so in their 3-D instantiation. Congruently with a 2-D-only surface code, the latter result is completely maintained if rich depth information is offered by the stimuli. Thus, the surface-conjunction effect with realistically rendered stimuli with stereoscopic depth and surface texture is just as great as, or greater than, if stimuli are impoverished with respect to depth information (without stereo or texture).

These observations are congruent with a surface-based shape representation in midlevel vision, such as described by Leek et al. (2005; for a related theory, see also Ullman, 2007; Ullman & Bart, 2004; Ullman et al., 2002; for implementations of surface-based models in computer vision, see Ashbrook et al., 1998; Barrow & Tenenbaum, 1981; Fan et al., 1989; Faugeras et al., 1983; Fisher, 1989; Lee & Park, 2002; Potmesil, 1983). According to this view, contours (regardless of the property that defines them) serve to segment an object into its visible surfaces, which are defined as 2-D edge-bounded polygons.

The notion of a surface-based code for shape representation may be implemented in the context of different theories of high-level visual shape perception. For instance, a surface code could be accommodated by image-based theories (e.g., Riesenhuber & Poggio, 2002; Tarr & Bülthoff, 1998), provided that they assume some degree of structure in the representation that is compatible with surfaces as one of its elementary units. At the opposite end of the theoretical spectrum, the present evidence may also be compatible with structural description theories. For instance, a commonly neglected aspect of Biederman's theory of recognition by components is its assumption that 2-D surfaces are part of the elementary units for shape representation, in addition to volumetric units (i.e., geons; Biederman, 1987; Hummel, 2001). If it is assumed that a 2-D surface-based code may be derived even for objects with a clear 3-D aspect which can also be represented in terms of volumetric primitives (as in the present experiments; see also Leek et al., 2005; Leek et al., 2008; Leek et al., 2018; Reppa et al., 2015; Reppa & Leek, 2019), then this theory would be congruent with the present findings.

More broadly speaking, given the insensitivity of the surface-based code demonstrated here, it is rather evident that it must be supplemented by some other mode(s) of shape representation capable of integrating depth information. Indeed, as noted in the Introduction, several demonstrations that depth information has a profound effect on shape perception have been reported (Aubin & Arguin, 2014; Bennett & Vuong, 2006; Blais et al., 2009; Burke, 2005; Burke et al., 2007; Cristino et al., 2015; Hong Liu et al., 2006; Lee & Saunders, 2011). Most relevant here, Blais et al. (2009) provided evidence that a particular stage of shape representation, which appears to code shape in terms of a structural description, can use depth information to render its representations invariant to both 2-D and 3-D rotations. This study used the visual-search task and assessed the performance costs occurring when the global shape properties of elongation and curvature that defined the target either were shared with distractors (conjunction effect) or had a value midway between those of distractors (linear nonseparability effect). Across several experiments, the results indicate that the shape representations underlying those conjunction and linear nonseparability effects are rotation-invariant provided that the depth information offered by the stimuli is sufficient.

Taken together, these investigations point to at least two distinct modes for shape representation (see also Foster & Gilson, 2002). We suggest that an important goal of future studies would be to further investigate these two representation modes together in the context of the same experiments to determine their concurrent use by the human visual system.

Conclusions

Our study of the shape conjunction effect in the visual-search task provides results in support of a 2-D surface-based code contributing to shape perception at an intermediate level of processing. Under this code, shapes are represented in terms of collections of 2-D edge-bounded polygons (i.e., surfaces). Considered in the larger context of the literature on shape perception, the present findings may be taken as evidence for one arm of a dual or multiple mode of shape representation in human vision.

Keywords: visual shape perception, depth information

Acknowledgments

We thank Aline Gauchat for her technical help in carrying out parts of the research reported in this article. This work benefited from the financial support of the Natural Sciences and Engineering Research Council of Canada in the form of a research grant to MA (Grant RGPIN-2014-03789).

Commercial relationships: none.
Corresponding author: Martin Arguin.
Email: martin.arguin@umontreal.ca.
Address: Centre de Recherche en Neuropsychologie Expérimentale et Cognition, Département de psychologie, Université de Montréal, Montréal, Canada.

References

Arguin, M., Bub, D., & Dudek, G. (1996). Shape integration for visual object recognition and its implication in category-specific visual agnosia. *Visual Cognition*, *3*, 221–275.

Arguin, M., & Saumier, D. (2000). Conjunction and linear non-separability effects in visual shape encoding. *Vision Research*, *40*, 3099–3115.

Ashbrook, A. P., Fisher, R. B., Robertson, C., & Werghi, N. (1998). Finding surface correspondence for object recognition and registration using pairwise geometric histograms. In H. Burkhardt & B. Neumann (Eds.), *Computer Vision—ECCV'98. ECCV 1998. Lecture Notes in Computer Science, Vol 1407* (pp. 674–686). Berlin, Heidelberg: Springer. https://doi.org/10.1007/BFb0054772.

Aubin, M., & Arguin, M. (2014). Stereo and shading contribute independently to shape convexity/concavity discrimination. *Perception*, *45*, 333–243.

Barr, A. H. (1981). Superquadrics and angle-preserving transformations. *IEEE Computer Graphics and Applications*, *1*, 1–11.

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

Bennett, D. J., & Vuong, Q. C. (2006). A stereo advantage in generalizing over changes in viewpoint on object recognition tasks. *Perception & Psychophysics*, *68*, 1082–1093, https://doi.org/10.3758/BF03193711.

Biederman, I. (1987). Recognition by components: A theory of human image understanding. *Psychological Review*, *94*, 65–96.

Blais, C., Arguin, M., & Marleau, I. (2009). Orientation invariance in visual shape perception. *Journal of Vision*, *9*(2):14, 1–23, https://doi.org/10.1167/9.2.14. [PubMed] [Article]

Blum, H. (1973). Biological shape and visual science (part I). *Journal of Theoretical Biology*, *38*, 205–287, https://doi.org/10.1016/0022-5193(73)90175-6.

Brooks, R. (1981). Symbolic reasoning among 3-dimensional and 2-dimensional images. *Artificial Intelligence*, *17*, 285–349.

Burbeck, C. A., & Pizer, S. M. (1995). Object representation by cores: Identifying and representing primitive spatial regions. *Vision Research*, *35*, 1917–1930, https://doi.org/10.1016/0042-6989(94)00286-U.

Burke, D. (2005). Combining disparate views of objects: Viewpoint costs are reduced by stereopsis. *Visual Cognition*, *12*, 705–719.

Burke, D., Taubert, J., & Higman, T. (2007). Are face representations viewpoint dependent? A stereo advantage for generalising across different views of faces, *Vision Research*, *47*, 2164–2169, https://doi.org/10.1016/j.visres.2007.04.018.

Carrasco, M., Evert, D. L., Chang, I., & Katz, S. M. (1995). The eccentricity effect: Target the eccentricity affects performance on conjunction searches. *Perception & Psychophysics*, *57*, 1241–1261.

Carrasco, M., Giordano, A. M., & McElree, B. (2006). Attention speeds processing across eccentricity: Feature and conjunction searches. *Vision Research*, *46*, 2028–2040.

Carrasco, M., & Yeshurun, Y. (1998). The contribution of covert attention to the set-size and eccentricity effects in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 673–692.

Cortese, J. M., & Dyre, B. P. (1996). Perceptual similarity of shapes generated from Fourier descriptors. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 133–143.

Cristino, F., Davitt, L., Hayward, W., & Leek, E. C. (2015). Viewpoint interpolation during object recognition is depth blind: Evidence from stereoscopic presentation. *Quarterly Journal of Experimental Psychology*, *68*, 2419–2436.

Crouzet, S., & Serre, T. (2011). What are the visual features underlying rapid object recognition? *Frontiers in Psychology*, *2*:326, https://doi.org/10.3389/fpsyg.2011.00326.

Delorme, A., Richard, G., & Fabre-Thorpe, M. (2010). Key visual features for rapid categorization of animals in natural scenes. *Frontiers in Psychology*, *1*:21, https://doi.org/10.3389/fpsyg.2010.00021.

Dugué, L., Xue, A. M., & Carrasco, M. (2017). Distinct perceptual rhythms for feature and conjunction searches. *Journal of Vision*, *17*(3):22, 1–15, https://doi.org/10.1167/17.3.22. [PubMed] [Article]

Duncan, J., & Humphreys, G. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458.

Duncan, J., & Humphreys, G. (1992). Beyond the search of surface: Visual search and attentional engagement. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 578–588.

Eckstein, M. P. (2001). The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. *Psychological Science*, *9*, 111–118.

Edelman, S., & Weinshall, D. (1991). A self-organizing multiple-view representation of 3D objects. *Biological Cybernetics*, *64*, 209–219.

Fan, T.-J., Medioni, G., & Nevatia, R. (1989). Recognizing 3-D objects using surface descriptions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*, 1140–1157.

Faugeras, O. D., Germain, F., Kryze, G., Boissonnat, J. D., Herbert, M., Ponce, J., . . . Ayache, N. (1983). Towards a flexible vision system. In A. Pugh (Ed.), *Robot Vision. International Trends in Manufacturing Technology* (pp. 129–142). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-09771-7_9.

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences, USA*, *103*, 18014–18019.

Fisher, R. (1989). *From surfaces to objects: Computer vision and three dimensional scene analysis*. Chichester, UK: John Wiley and Sons.

Foster, D. H., & Gilson, S. J. (2002). Recognizing novel three-dimensional objects by summing signals from parts and views. *Proceedings of the Royal Society of London B: Biological Sciences*, *269*, 1939–1947.

Hong Liu, C., Ward, J., & Young, A. W. (2006). Transfer between two-and three-dimensional representations of faces. *Visual Cognition*, *13*, 51–64, https://doi.org/10.1080/13506280500143391.

Hummel, J. E. (2001). Complementary solutions to the binding problem in vision: Implications for shape perception and object recognition. *Visual Cognition*, *8*, 489–517.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, *99*, 480–517.

Hung, C. C., Carlson, E. T., & Connor, C. E. (2012). Medial axis shape coding in macaque inferotemporal cortex. *Neuron*, *74*, 1099–1113, https://doi.org/10.1016/j.neuron.2012.04.029.

Kimia, B. B. (2003). On the role of medial geometry in human vision. *Journal of Physiology–Paris*, *97*, 155–190, https://doi.org/10.1016/j.jphysparis.2003.09.003.

Lades, M., Vorbruggen, J. C., Buhmann, J., Lange, J., von der Marlsburg, C., Wurtz, R. P., & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, *42*, 300–311.

Lee, Y.-L., & Park, R.-H. (2002). A surface-based approach to 3-D object recognition using a mean field annealing neural network. *Pattern Recognition*, *35*, 299–316.

Lee, Y. L., & Saunders, J. A. (2011). Stereo improves 3D shape discrimination even when rich monocular shape cues are available. *Journal of Vision*, *11*(9):6, 1–12, https://doi.org/10.1167/11.9.6. [PubMed] [Article]

Leek, E. C., Reppa, I., & Arguin, M. (2005). The structure of 3D object shape representations: Evidence from whole-part matching. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 668–684.

Leek, E. C., Reppa, I., Rodriguez, E., & Arguin, M. (2008). Surface versus volumetric structure in the representation of 3-D object shape. *Quarterly Journal of Experimental Psychology*, *62*, 814–829.

Leek, E. C., Roberts, M. V., Dundon, N. M., & Pegna, A. J. (2018). Early sensitivity of evoked potentials to surface and volumetric structure during the visual perception of three-dimensional object shape. *European Journal of Neuroscience*. Advance online publication. https://doi.org/10.1111/ejn.14270.

Marr, D. (1982). *Vision*. New York: Freeman.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three dimensional shapes. *Proceedings of the Royal Society of London*, *207*, 187–217.

McElree, B., & Carrasco, M. (1999). The temporal dynamics of visual search: Evidence for parallel processing in future and conjunction searches. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1517–1539.

Nakayama, K., He, Z., & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In S. M. Kosslyn & D. Osherson (Eds.), *Visual cognition: Vol 2. An invitation to cognitive science* (2nd ed., pp 1–70). Cambridge, MA: MIT Press.

Nakayama, K., & Shimojo, S. (1992, September 4). Experiencing and perceiving visual surfaces. *Science*, *257*(5075), 1357–1363.

Nefs, H. T. (2008). Three-dimensional object shape from shading and contour disparities. *Journal of Vision*, *8*(11):11, 1–16, https://doi.org/10.1167/8.11.11. [PubMed] [Article]

Norman, J. F., & Todd, J. T. (1996). The discriminability of local surface structure. *Perception*, *25*, 381–398.

Norman, J. F., & Todd, J. T. (1998). Stereoscopic discrimination of interval and ordinal depth relations on smooth surfaces and in empty space. *Perception*, *27*, 257–272.

Norman, J. F., Todd, J. T., & Phillips, F. (1995). The perception of surface orientation from multiple sources of optical information. *Perception & Psychophysics*, *57*, 629–636.

Norman, J. F., Todd, J. T., Norman, H. F., Clayton, A. M., & McBride, T. R. (2006). Visual discrimination of local surface structure: Slant, tilt, and curvedness. *Vision Research*, *46*, 1057–1069.

Oliver, Z. J., Cristino, F., Roberts, M. V., Pegna, A. J., & Leek, E. C. (2018). Stereo viewing modulates three-dimensional shape processing during object recognition: A high-density ERP study. *Journal of Experimental Psychology: Human Perception and Performance*, *44*, 518–534.

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of the visual search. *Vision Research*, *40*, 1227–1268.

Pasqualotto, A., & Hayward, W. G. (2009). A stereo disadvantage for recognising rotated familiar objects. *Psychonomic Bulletin & Review*, *16*, 832–838.

Pentland, A. P. (1986). Perceptual organization and the representation of form. *Artificial Intelligence*, *28*, 292–331.

Pizlo, Z. (2008). *3D shape: Its unique place in visual perception*. London: MIT Press.

Poggio, T., & Edelman, S. (1990, January 18). A network that learns to recognize 3-dimensional objects. *Nature*, *343*, 263–266.

Potmesil, M. (1983). Generating models of solid objects by matching 3D surface segments. In A. Bundy (Ed.), *Proceedings of the 8th International Joint Conference on Artificial Intelligence* (pp. 1089–1093). Karlsruhe, Germany: William Kaufmann.

Reppa, I., Greville, W. J., & Leek, E. C. (2015). The role of surface-based representations of shape in visual object recognition. *Quarterly Journal of Experimental Psychology*, *68*, 2351–2369, https://doi.org/10.1080/17470218.2015.1014379.

Reppa, I., & Leek, E. C. (2019). Surface diagnosticity predicts the high-level representation of regular and irregular object shape in human vision. *Attention, Perception & Psychophysics*, *81*(5), 1589–1608, https://doi.org/10.3758/s13414-019-01698-4.

Riesenhuber, M., & Poggio, T. (2002). Neural mechanisms of object recognition. *Current Opinion in Neurobiology*, *12*, 162–168.

Saumier, D., & Arguin, M. (2003). Distinct mechanisms account for the linear non-separability and conjunction effects in visual shape encoding. *Quarterly Journal of Experimental Psychology*, *56A*, 1373–1388.

Saunders, J. A., & Backus, B. T. (2006). Perception of surface slant from oriented textures. *Journal of Vision*, *6*(9):3, 882–897, https://doi.org/10.1167/6.9.3. [PubMed] [Article]

Stankiewicz, B. J. (2002). Empirical evidence for independent dimensions in the visual representation of three-dimensional shape. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 913–932.

Tarr, M. J., & Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition*, *67*, 1–20.

Todd, J. T., & Norman, J. F. (2003). The visual perception of 3-D shape from multiple cues: Are observers capable of perceiving metric structure? *Perception and Psychophysics*, *65*, 31–47.

Todd, J. T., Norman, J. F., Koenderink, J. J., & Kappers, A. M. (1997). Effects of texture, illumination and surface reflectance on stereoscopic shape perception. *Perception*, *26*, 807–822.

Townsend, J. T. (1972). Some results concerning the identifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology*, *25*, 168–199.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Treisman, A. M., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, *95*, 15–48.

Treisman, A. M., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 459–478.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, *32*, 193–254.

Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences*, *11*, 58–64, https://doi.org/10.1016/j.tics.2006.11.009.

Ullman, S., Assif, L., Fetaya, E., & Harari, D. (2016). Atoms of recognition in human and computer vision. *Proceedings of the National Academy of Sciences, USA*, *113*, 2744–2749.

Ullman, S., & Bart, E. (2004). Recognition invariance obtained by extended and invariant features. *Neural Networks*, *17*, 833–848, https://doi.org/10.1016/j.neunet.2004.01.006.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, *5*, 682–687, https://doi.org/10.1038/nn870.