Automatic Feature-Based Grouping During Multiple Object Tracking

Gennady Erlikhman, Brian P. Keane, Everett Mettler, Todd S. Horowitz, and Philip J. Kellman

Automatic Feature-Based Grouping During Multiple Object Tracking

Gennady Erlikhman
University of California, Los Angeles

Brian P. Keane
Rutgers University and UMDNJ–Robert Wood Johnson Medical
School

Everett Mettler
University of California, Los Angeles

Todd S. Horowitz
Brigham and Women's Hospital, Boston, Massachusetts, and
Harvard Medical School

Philip J. Kellman
University of California, Los Angeles

Contour interpolation automatically binds targets with distractors to impair multiple object tracking (Keane, Mettler, Tsoi, & Kellman, 2011). Is interpolation special in this regard or can other features produce the same effect? To address this question, we examined the influence of eight features on tracking: color, contrast polarity, orientation, size, shape, depth, interpolation, and a combination (shape, color, size). In each case, subjects tracked 4 of 8 objects that began as undifferentiated shapes, changed features as motion began (to enable grouping), and returned to their undifferentiated states before halting. We found that intertarget grouping improved performance for all feature types except orientation and interpolation (Experiment 1 and Experiment 2). Most importantly, target-distractor grouping impaired performance for color, size, shape, combination, and interpolation. The impairments were, at times, large (>15% decrement in accuracy) and occurred relative to a homogeneous condition in which all objects had the same features at each moment of a trial (Experiment 2), and relative to a "diversity" condition in which targets and distractors had different features at each moment (Experiment 3). We conclude that feature-based grouping occurs for a variety of features besides interpolation, even when irrelevant to task instructions and contrary to the task demands, suggesting that interpolation is not unique in promoting automatic grouping in tracking tasks. Our results also imply that various kinds of features are encoded automatically and in parallel during tracking.

Keywords: multiple object tracking, attention, perceptual grouping, perceptual organization, illusory contours

Supplemental materials: http://dx.doi.org/10.1037/a0031750.supp

If you throw a handful of marbles on the floor, you will find it difficult to view at once more than six, or seven at most, without confusion; but if you group them into twos, or threes, or fives, you can comprehend as many groups as you can units; because the mind considers these groups only as units—it views them as wholes, and throws their parts out of consideration.

—William Hamilton, Lectures on Metaphysics, Lecture XIV (quoted by William James, p. 406, *Principles of Psychology*)

Since Duncan's (1984) seminal demonstration, it has been widely agreed that attention can be directed to objects. However, a satisfying recipe for visual objecthood is still elusive. A major contribution came from Scholl, Pylyshyn, and Feldman (2001), who made two important points. First, they argued that we cannot take for granted that the properties that lead to perceptual grouping (e.g., Gestalt laws) will necessarily be the properties that form the units of attention. Second, they proposed a method to empirically test the properties that define objects, using the multiple object tracking paradigm (MOT, Pylyshyn & Storm, 1988).

In the classic MOT experiment, subjects are presented with an array of identical objects, a subset of which are designated as "targets," and then are asked to track those targets as they move independently and unpredictably for several seconds. A subject's ability to correctly identify the targets at the end of this period of motion quantifies the ability to maintain an object representation over time and space. One of the most interesting findings was that viewers of these displays could track the targets but could not keep track of object identities (Pylyshyn, 2004). For example, if each target was assigned a name at the start of a trial, subjects could pick out the targets from all of the objects at the end of the trial but could not remember which target had which name. This was attributed to greater confusability among the targets due to inhib-

itory effects on nontargets (Pylyshyn, 2006). It was suggested that features of nontargets were not processed to the same extent and that grouping would not apply to nontargets.

In Scholl et al.'s (2001) adaptation, each target was connected to a distractor via luminance-defined lines. The authors found that when the connecting line did not physically touch the targets' borders, there was very little effect on tracking. However, when the line touched each object, it was much more difficult to correctly report the targets, indicating that this manipulation united two independently moving stimuli into a single attentional unit or object. The disrupting effect of the connecting line could be described as falling under the Gestalt principle of uniform connectedness (Palmer & Rock, 1994).

Recently, Keane et al. (2011) demonstrated that targets and distractors could be linked by contour interpolation, which does not involve any physical connection between the items. Keane et al. used a variant of MOT dubbed "multiple vertex tracking" (MVT, see also Franconeri, Jonathan, & Scimeca, 2010; Howe, Cohen, Pinto, & Horowitz, 2010; Mettler, Keane, & Kellman, 2008). In MVT, the items do not move independently around the screen; instead, they are grouped into four doublets, one in each quadrant of the screen. One item in each doublet is a target and the other is a distractor. Each quadrant pair orbits its respective barycenter (see Figure 1, and online supplemental materials for movies). As movement begins, a sector of each disk is removed to create "Pac-Man" figures that produce illusory quadrilaterals. One Pac-Man from each quadrant forms one corner of one quadrilateral and the remaining "Pac-Men" form the corners of another. These quadrilaterals morph in shape as the Pac-Men orbit in their respective quadrants. This setup allows objects to be large and nonoverlapping and therefore ensures high contour salience (Shipley & Kellman, 1992). The contours occasionally connected targets with one another to potentially aid tracking, presumably through reducing the effective tracking load; in other conditions, the contours connected targets with distractors, making the two kinds of objects harder to distinguish (Scholl et al., 2001). Keane et al. obtained two primary findings. First, when the four targets formed one illusory quadrilateral and four distractors formed the other, tracking accuracy improved modestly relative to when there was no interpolation. Second, and more importantly, when two targets and two distractors formed each of the two quadrilaterals, accuracy worsened relative to when interpolation was absent. The latter effect could be modulated in a stimulus-driven fashion: Decreasing

the support ratio (percentage of the contour that was physically specified) or increasing the relative rotation angle of the Pac-Men reduced the contour strength and the impact of the contours on tracking. Such results suggest that (a) interpolation is "automatic" in that it occurs without explicit instruction and contrary to the task demands, (b) attentional allocation can serve as an index for earlier perceptual organization processes, and (c) multiple features are utilized during tracking, at least when those features are relevant to interpolation (e.g., size, orientation).

In light of these findings, Keane et al. (2011) proposed the *segmentation-relevance principle*, according to which featural properties affect the formation of attentional units in tracking only when such features guide scene segmentation. This principle specifically predicts that properties such as color should *not* produce automatic grouping during MOT. The aim of this article is to test the predictions of the segmentation-relevance principle.

The segmentation-relevance principle builds on a long tradition of research showing that contour interpolation is ontogenetically, phylogentically, and (in some cases) physiologically early (Kellman & Spelke, 1983; Nieder, 2002; Valenza & Bulf, 2011; von der Heydt, Peterhans, & Baumgartner, 1984), and that it plays a crucial role for determining the number, shape, and persistence of objects in cluttered distal environments. Additionally, interpolation not only *groups* separated items, it also represents—phenomenally and functionally—the regions contained between the items grouped. More simply, interpolation leads to "filling-in," which itself has functional consequences on behavior over and above the consequences that arise from grouping without filling-in (Davis & Driver, 1998; Gold, Murray, Bennett, & Sekuler, 2000; He & Nakayama, 1992; Moore, Yantis, & Vaughan, 1998; Rensink & Enns, 1998).Therefore, interpolation may very well be unique in its capacity to automatically direct attention during tracking.

Accordingly, there is substantial empirical evidence that featural properties irrelevant to scene parsing are also irrelevant to MOT, especially when the task instructions do not explicitly ask the subject to attend to or use the features or when the usage of those features cannot be used to improve tracking (Feldman & Tremoulet, 2006; Flombaum, Scholl, & Santos, 2009; Mitroff & Alvarez, 2007; Pylyshyn, 2004; Scholl, 2007, 2009). In a study by Makovski and Jiang (2009b), when two targets and two distractors shared one color and the remaining objects shared a different color (to enable target-distractor grouping), there was not a consistent

Targets Distractors Group (TDG)
Experiments 1, 2 & 3

Targets Group (TG)
Experiment 1

Homogenous
Experiment 2

Diversity
Experiment 3
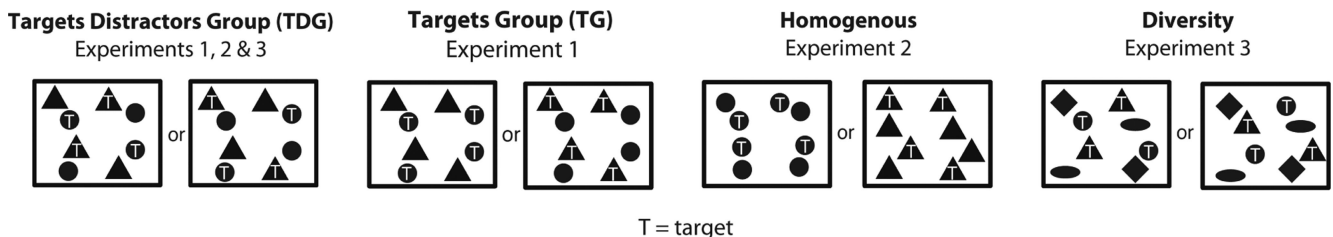
T = target

Figure 1. Distribution of feature tokens across targets and distractors for the shape feature type in each of the three experiments. In the TDG condition, two targets and two distractors possessed one feature token and the remaining four objects had the remaining feature token. In the TG condition, one feature token was assigned to the targets and the other to the distractors. In the homogeneous and diversity conditions, there was no target-distractor grouping because either all objects had the same shape, or the targets and distractors did not share a shape. Features were distributed in an analogous fashion for the other feature types (see Figure 2).

detriment relative to a homogeneous condition in which all objects shared the same color. Viswanathan and Mingolla (2002) found a similar null result for color and also for shape (squares or hashes). Viswanathan and Mingolla did, however, find an advantage of depth: When targets and distractors were equally distributed between two depth planes, subjects had better tracking performance than when all objects shared the same plane. The authors argued that the effect was due not to grouping but to the spreading and segregation of attention to surfaces; that is, there were fewer target-distractor pairs to track per surface and this led to improved performance. For our purposes, the salient point is that spatial properties like depth are useful for scene parsing, whereas featural properties like color are not.

A few studies have shown that there can be grouping on the basis of spatiotemporal properties (speed, direction, location). Suganuma and Yokosawa (2006), for example, found that when targets and distractors shared a chasing relationship (common motion) in their motion patterns, tracking performance was worse than when no such relationship existed. Yantis (1992) showed that targets can be segregated from distractors on the basis of their velocities, even when the two velocities could not be reliably discriminated, and that this segregation improves tracking relative to when all objects move at similar velocities. Because spatiotemporal properties are those that are considered to be essential for "visual objecthood," it is perhaps not surprising that they should have an important role in tracking (Scholl, 2009). Here, we examine whether the *featural* properties of objects—that is, the features that define the shape or surface of objects—are automatically used for grouping.

There is evidence that subjects track on the basis of featural properties when relevant to the task (Cohen, Pinto, Howe, & Horowitz, 2011; Horowitz et al., 2007; Howard & Holcombe, 2008; Oksama & Hyönä, 2004, 2008; Pinto, Howe, Cohen, & Horowitz, 2010). Even when feature information is irrelevant, it can be used to *improve* tracking performance. Horowitz et al. (2007), using cartoon animal stimuli, found that tracking performance improved when all stimuli were unique, relative to a homogeneous condition. However, when each target was unique, but paired with an identical distractor (potentially allowing for target-distractor grouping), the advantage was eliminated, though performance was not significantly worse than in the homogenous condition. Makovski and Jiang (2009a) obtained similar findings using a more controlled stimulus set of colored digits. When all digits were unique in color or identity, performance improved in comparison with a homogenous condition. However, subjects could not use unique conjunctions of features to improve tracking performance, indicating that the effects on tracking were due to the object features, not their uniqueness. Likewise, Störmer, Li, Heelkeren, and Lindenberger (2011) showed that targets were easier to track when they were colored differently from distractors. These studies, considered as a whole, suggest that task-irrelevant features are used in tracking, but only when helpful to the task. To our knowledge, no study has convincingly demonstrated the automatic usage of (noninterpolation) featural properties during MOT, much less automatic feature-based grouping.

In the present article, the segmentation-relevance principle was tested with an MVT paradigm. Over the course of three experiments, we examined eight candidate feature types: color, contrast polarity, orientation, size, shape, a combination (of color, shape,

and size), stereo depth, and interpolation (see Figure 2). These were selected because they are well known to guide perceptual organization (contrast, Chan & Hayward, 2009; Earle, 1999; luminance, Gilchrist, Humphreys, Riddoch, & Neumann, 1997; Sekuler & Bennett, 2001; size, Gori & Spillmann, 2010; color, Pessoa, Beck, & Mingolla, 1996; Rock, Nijhawan, Palmer, & Tudor, 1992; shape, Hadad & Kimchi, 2008; Rock, Nijhawan, Palmer, & Tudor, 1992; Vickery & Jiang, 2009; orientation: Beck, 1966; Nothdurft (1985); depth: Nakayama, Shimojo, & Silverman, 1989; Viswanathan & Mingolla, 2002). The aforementioned features also likely guide the deployment of attention in visual search (Wolfe & Horowitz, 2004), indicating that they could be important for MOT.

In all three experiments, we compared the critical *target-distractor grouping* (TDG) condition with a different control condition. In the TDG condition, two targets and two distractors shared one feature, and the remaining objects shared the other feature (see Figure 2). The TDG condition was designed so that feature information would work *against* the subject by grouping targets with distractors while segregating targets from one another. In Experiment 1, we compared the TDG condition with a *target group* (TG) condition, where targets all shared one feature and all distractors shared another (see Figure 2). In the TG condition, feature information was expected to work *with* the subject, segregating targets from distractors and grouping targets together. In Experiment 2, the TDG condition was compared with a *homogeneous* condition, in which all objects always shared the same features (see Figure 2). Finally, in Experiment 3, the TDG condition was compared with a *diversity* condition, in which two targets shared one feature and the remaining targets shared another (as in the TDG condition), while the distractors were similarly divided between two additional features (to prohibit target-distractor grouping). In all three experiments, grouping strength was operationalized as the amount by which tracking performance in the TDG condition was eclipsed by the control condition (TG, homogenous, or diversity). If TDG causes an impairment in each of the three experiments for feature types besides interpolation, then the segmentation relevance principle will be disconfirmed.

Experiment 1: Which Features Can Be Used for Grouping During Tracking?

The purpose of Experiment 1 was to determine the sorts of features that are spontaneously used in tracking, which is an important first step in identifying those that induce automatic grouping. The key comparison for each feature type was between TG and TDG conditions. In the TG condition, all of the targets shared a single feature, whereas all of the distractors shared the opposing feature. In the TDG, two targets and two distractors shared one feature and the remaining objects shared the opposing feature. If a grouping process automatically guided attention, we would expect to see better tracking performance (higher accuracy) in the TG condition relative to the TDG condition.

Importantly, we wanted to test whether these features had an influence on grouping *during* tracking, not whether these features could be remembered and used for target selection. If target features were present throughout the entire trial, then
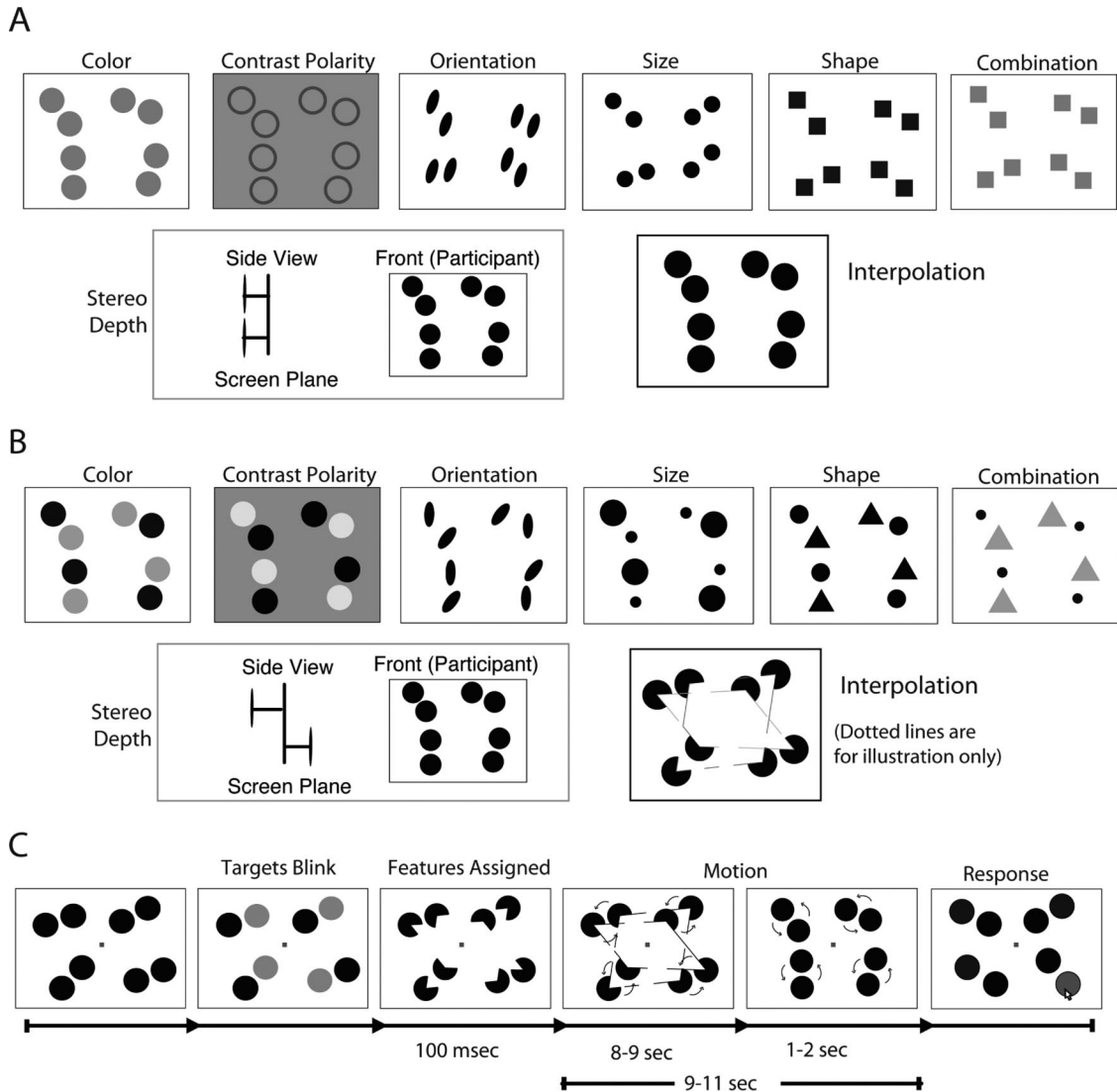
Figure 2. Stimuli and trial sequence for Experiment 1 and 2. (A) Starting and ending appearance of objects for each feature type. At the beginning and end of a trial, objects were undifferentiated and possessed features that were different from those that appeared during the motion phase. Starting colors for the color and combination conditions were yellow. See online supplemental materials for a color version. (B) Possible screenshots of the eight feature types during the motion phase. These screenshots could derive from either the TG or the TDG grouping relations. In all cases, a quadrant contained a single target and distractor. Color changes were to red and green in the color and combination conditions. (C) Phases of a trial for the interpolation condition. After target designation, objects (on certain conditions) became distinctive so that grouping could potentially occur. Feature changes (third panel) were instantaneous. Objects remained stationary for 100 ms after the initial feature change. Objects orbited around a central point in each quadrant and then returned to their initial, undifferentiated states a moment before halting.

subjects would have no need to track. For example, if all of the targets remained red and all of the distractors remained green throughout a trial, subjects could simply remember that targets were red, and then select the red items at the end of the trial. Therefore, all objects in our study began as identical objects during the stationary target-designation phase and simultaneously transformed with new features as the motion commenced. The objects returned to their initial, identical state at a randomly

selected point before the end of the motion phase. If features are employed during tracking, their employment must occur during tracking itself.

Method

Subjects. Twenty-one undergraduate students from the University of California, Los Angeles (UCLA) participated for

course credit. Subjects reported normal or corrected-to-normal vision. All were given a RanDot stereoacuity test (Stereo Optical Co., Inc.) and had a stereoacuity of at least 70 seconds of arc at 40.64 cm.

Equipment. All displays were in programmed in Matlab with Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Stimuli were displayed on a ViewSonic G250 CRT monitor, which was powered by a Macintosh PowerPC with a 2.1 GHz processor. Stereoscopic depth was created by using the CrystalEyes 3D goggles. Displays were shown at a resolution of 1024×768 pixels with a refresh rate of 120 Hz (60 Hz per eye). Subjects used a mouse and keyboard to register responses.

Stimuli. The fixation point was a blue square (0.4 degrees on a side). Displays were made of four pairs of objects, one in each screen quadrant. Objects within a quadrant were separated by 3.8 degrees and orbited a central point (barycenter) without rotating. The four barycenters were positioned on the corners of a 7.83×7.67 degree rectangle. Angular speeds were randomly assigned to a quadrant without replacement from the following values: 1.25, 1.67, 2.08, or 2.51 π rad/s. The direction of rotation (clockwise or counterclockwise) was randomly chosen for each. The angular velocity varied between not within quadrants. During motion, objects instantaneously transformed according to one of eight feature types, which are listed below (see Figure 2).

For each feature type, there were two grouping relations. In the TG condition, all of the targets transformed to having one feature token, whereas all of the distractors transformed to having a different feature token. For example, for the color feature type, all of the target circles became red and all of the distractors turned green (or vice versa). In the TDG condition, two distractors and two targets underwent the same transformation (e.g., all red), and two distractors and two targets underwent the other transformation (e.g., all green).

Color. Objects began and ended as yellow circles (radius = 1.84 degrees) on a white background and changed to red or green. During target designation, target circles flashed gray.

Contrast polarity. Objects began and ended as dark gray rings (radius = 1.84 degrees) on a light gray background and transformed into filled-in circles that were either white or black. During target designation, target rings flashed black.

Orientation. Objects began and ended as black ovals (major axis = 1.84 degrees, minor axis = 0.92 degrees) oriented at 67.5 degrees on a white background. They transformed into ovals oriented at 90 degrees or 45 degrees. Similar orientation differences altered tracking in the interpolation study (Keane et al., 2011, Experiment 4), and therefore this magnitude was thought to be sufficient for this experiment. During target designation, target ovals flashed gray.

Size. Objects began and ended as black circles (radius = 1.56 degrees) on a white background. They transformed into black circles with radii of 1.84 or 1.04 degrees. During target designation, target circles flashed gray.

Shape. Objects began and ended as black squares (side = 2.6 degrees) on a white background and transformed into either equilateral triangles (side = 1.43 degrees) or circles (radius = 1.3 degrees). During target designation, target squares flashed gray.

Combination. Objects began and ended as yellow squares (side = 2.17 degrees) on a white background. Four of the squares transformed into circles and four into triangles. Half of the objects were red and half were green. Half of the objects became small and half became large (as in the size condition [small radius/side = 1.04 degrees, large radius/side = 1.84 degrees]; see Procedure for how the colors, shapes, and sizes were combined). During target designation, target squares flashed gray.

Stereo depth. Objects began and ended as black circles (radius = 1.56 degrees) on a white background. A disparity was added that specified objects on a frontoparallel plane 1.46 cm in front of the screen. Depth was introduced by shifting each eye's image. Equal lateral shifts produced a virtual depth of 13.13 cm in front of the screen for the cross disparity stimuli and 17.06 cm behind the screen for the uncrossed disparity stimuli. During target designation, circles flashed gray.

Interpolation. Objects began as black circles (radius = 1.56 degrees) on a white background, with one sector of each circle removed to produce Pac-Men. Groups of four were relatable (Kellman & Shipley, 1991), producing a dynamically morphing illusory quadrilateral. During target designation, circles flashed gray.

Procedure. Subjects sat 114 cm away from the monitor so that each pixel subtended .02 degrees of visual angle. Subjects were shown instructions on the screen explaining the tracking task. They were informed that the objects that they were tracking would occasionally change in appearance and that, despite the changes, they should track them continuously. No mention was made of grouping or illusory figures in the instructions or in the recruitment phase of the experiment. At the start of each trial, a blue fixation square and four pairs of objects were displayed, one pair per quadrant. The targets (one per pair) flashed for 200 ms, all objects transformed their features instantaneously (to enable grouping), and after a 100-ms delay, each quadrant pair began orbiting a point (barycenter) centered within that quadrant. The 100-ms delay was meant to prevent the "flash jump" effect, in which objects appear to jump ahead when simultaneously beginning motion and changing appearance (Eagleman & Sejnowski, 2007; Keane et al., 2011). Objects maintained their orientation throughout the motion phase. After orbiting for 8 s, objects transformed back into their initial identical appearance at a random point within the following second, continued moving for another second, and finally stopped at a random point within the following second. Therefore, the total motion duration on a given trial randomly varied between 9 and 11 s (see Figure 2C). This design follows what others have done and is thought to discourage strategies of comparing the first and last frames of the motion (Cohen et al., 2011; Horowitz et al., 2007; Makovski & Jiang, 2009a; Oksama & Hyönä, 2004, 2008; Pinto et al., 2010). Subjects were given an unlimited amount of time to select all four targets with the mouse. Selected circles were highlighted by the same luminance change that defined them as targets in the target designation phase. After selection, subjects were informed of the number of correctly identified targets on that trial and their cumulative percent correct. Subjects were not informed of which of the four responses were incorrect.

The experiment consisted of seven blocks. Each block was comprised of one trial of each grouping relation (TG vs. TDG) for each of the eight feature types, for a total of 16 trials. Trial order was randomized within each block. Subjects were instructed to take a break between blocks. For the TDG conditions, the pairing between features and location was randomized. For example, for color trials, the target in the upper left quadrant turned red about

half the time, and green the other half. For the TG conditions, features were quasi-randomly assigned to the targets. For example, in the color TG trials, half of the time the targets were all red, and half of time the targets were all green. Each experiment incorporated only seven TG trials for each feature type. Because there were two kinds of TG trials (e.g., red and green), some subjects received three trials of one alternative (e.g., red) and four of the other (e.g., green). The feature alternative that was used more frequently was counterbalanced across subjects. The entire experiment lasted about 50 min.

At the start of the experiment, subjects had three practice trials, which always involved eight identical black disks (no feature changes). In all other respects, the practice trials were the same as the color nonpractice trials and were excluded from the analysis. After completing the experiment, subjects were given a paper-based questionnaire with the following question: "What strategy or strategies did you use to track objects in the experiment?"

Results and Discussion

Results are shown in Figure 3. To identify the feature types that influenced grouping, we performed (two-tailed) planned comparisons of the two grouping relations—TG and TDG—for each feature type. The grouping relation altered performance for seven of the eight feature types: color, $t(20) = 6.943$, $p < .0001$, Cohen's d: 1.623; contrast polarity, $t(20) = 4.579$, $p < .001$, Cohen's d: 1.025; size, $t(20) = 5.657$, $p < .0001$, Cohen's d: 1.129; shape, $t(20) = 4.639$, $p < .001$, Cohen's d: 0.988; depth, $t(20) = 4.917$, $p < .0001$, Cohen's d: 1.311; combination, $t(20) = 7.747$, $p < .0001$, Cohen's d: 1.720; and interpolation, $t(20) = 3.867$, $p < .001$, Cohen's d: 0.683. Orientation did not quite elicit a significant grouping effect, $t(20) = 1.910$, $p = .071$, Cohen's d: 0.390.

There are three notable findings here. First, and most importantly, besides interpolation, subjects utilized a range of features to group objects during tracking and this occurred without any explicit instructions. The magnitude of the TG/TDG differences was large, sometimes exceeding 20% (e.g., color). Second, the magni-

tude of the TG/TDG difference for the combination features did not surpass that of the individual feature types that composed it (size, color and shape). If feature types do combine or interact to promote grouping, it is in a highly nonlinear way. Finally, common orientation did not affect performance, even though orientation changes of 48 degrees were critical for the interpolation effect (Keane et al., 2011). This is an interesting example where feature relations (viz, relatability; Kellman & Shipley, 1991), but not common features, may be more important for grouping. It should be noted, however, that common orientation may cause automatic grouping in other cases, such as when the objects are more slender or when the orientation differences are more dramatic (see Discussion).

Experiment 2: Does Feature-Based Grouping Impair Tracking?

The first experiment showed that objects can be grouped based on a variety of features, even when the features in question are task-irrelevant. However, Experiment 1 leaves open the important question of whether the grouping effects reflect facilitation in the TD condition, impairment in the TDG condition, or both. If they arise from pure facilitation, then our results extend the findings of Horowitz et al. (2007) and Makovski and Jiang (2009b) without providing strong support of automaticity. Facilitation could arise, for example, from subjects noticing the featural similarity of the targets and deciding to track those targets on that basis. To add stronger evidence for automaticity, we compared the TDG condition with a *homogenous* control condition in which all objects shared the same features. Lower performance in TDG trials relative to homogenous trials would mean that the effects in Experiment 1 could not be ascribed (only) to facilitation or voluntary grouping strategies; they must also arise from automatic feature based grouping.
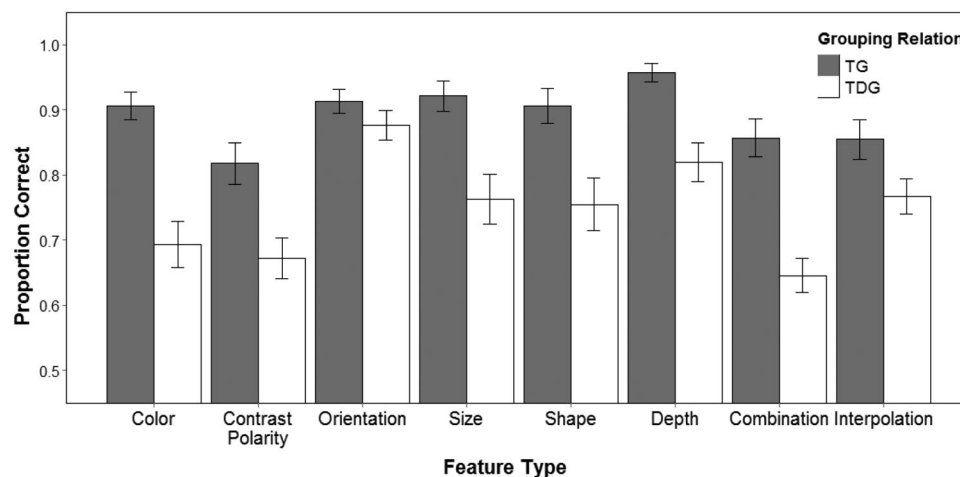

Figure 3. Average proportion correct in Experiment 1 as a function of feature type (x-axis) and grouping relation (Targets-Group [TG]; and Targets-Distractors Group [TDG]). The TG/TDG differences were significant for all feature types except orientation.

Method

Subjects. Twenty-three UCLA undergraduates participated for course credit. All reported normal or corrected-to-normal vision, passed the same stereoacuity test as in Experiment 1, and did not participate in Experiment 1.

Equipment and stimuli. The equipment was the same as that for Experiment 1. There were two grouping relations. The TDG condition was identical to that in Experiment 1. In the homogenous condition, objects transformed to exactly one feature token. For example, in the homogenous shape condition, all objects began as squares, transformed to all triangles or all circles, and then transformed back to squares (see Figures 1 and 2). In the homogenous interpolation condition, all objects began and ended as circles, as before, but changed to identical Pac-Men during the motion phase. The sectors of the Pac-Men were the same as those of a single randomly chosen Pac-Man from the TDG condition (so that all objects were homogeneous at each moment). Because both the TDG and homogenous conditions involved the same abrupt feature changes, any differences between the two can be attributed to grouping.

Procedure. The procedure was similar to that of Experiment 1. Half of the homogenous trials used only one of the features and half used only the other. Because there were seven blocks with one homogenous trial for each feature per block, there was an unequal number of these features. For example, across homogenous color trials in a single experimental run, all stimuli were red in three trials and green in four trials, or vice versa. Whether a subject received three red or three green color trials was randomized across subjects. The same randomization applied to all other feature types. Each block contained one trial of each feature type and grouping relation. The order of trials within each block was randomized as in Experiment 1.

Results and Discussion

As before, two-tailed *t* tests were performed for each feature type. The effect of feature type was significant for: color, $t(22) =$ 5.874, $p < .0001$, Cohen's *d*: 1.101; size, $t(22) = 4.034$, $p < .001$, Cohen's *d*: 0.939; shape, $t(22) = 4.522$, $p < .001$, Cohen's *d*: 0.627; combination, $t(22) = 5.558$, $p < .0001$, Cohen's *d*: 1.001; and interpolation, $t(22) = 3.165$, $p = .0045$, Cohen's *d*: 0.523. The grouping relation was irrelevant for: contrast polarity, $t(22) =$ 1.290, $p = .210$, Cohen's *d*: 0.321; orientation, $t(22) = -0.371$, $p = .715$, Cohen's *d*: 0.0542; and depth, $t(22) = .0725$, $p = .476$, Cohen's *d*: 0.113.

Although not our primary aim, we also considered whether there were facilitation effects. First, the TDG conditions of the two experiments were compared to verify that the two experiments were comparable. It was found that for each feature type, the TDG conditions were statistically equivalent (uncorrected *t* tests; all ps > 0.15). Next, the TDG–TG difference in the first experiment was compared with the TDG–homogenous difference in the second experiment. The second difference turned out to be smaller for contrast polarity, size, depth, and combination (all ps$< .05$); the difference was marginally smaller for color ($p =$ .07) and shape ($p = .055$). This suggests that—except for interpolation and orientation—intertarget grouping produced at least marginally significant facilitation for each feature type (Figure 4). We speculate that the grouping factors that elicited an advantage in Experiment 1 without a detriment in Experiment 2 (contrast polarity and depth) afforded the use of an explicit grouping strategy without automatic grouping.

To summarize, by comparing TDG performance to a homogenous baseline, we found evidence for automatic, feature-based grouping. In the color, size, shape, combination, and interpolation conditions, performance was significantly reduced when two targets and two distractors formed one dynamically morphing group and the remaining four objects formed another. This grouping was automatic in that it occurred even though it was consistently detrimental to task performance and irrelevant to task instructions. Other features, like depth and contrast polarity, improved tracking without hindering it. Orientation differences, meanwhile, did not lead to any appreciable grouping effect either way.
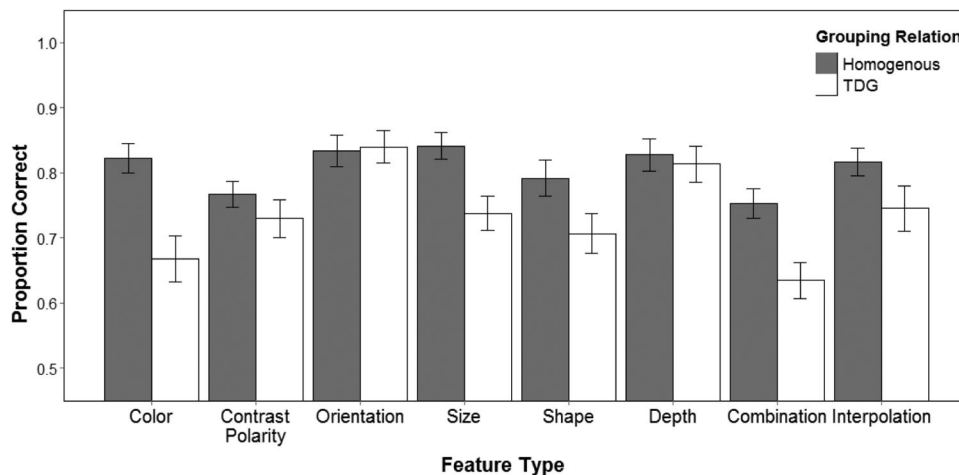

Figure 4. Average tracking performance as a function of feature type and grouping relation in Experiment 2. When targets and distractors grouped (TDG), tracking accuracy was lower versus when all objects were homogeneous. This was the case for all feature types except contrast, orientation, and depth.

Experiment 3: Controlling for Target Feature Diversity

In the first two experiments, it is possible that subjects performed worse in the TDG condition not because of target-distractor grouping, but because the targets in that condition had a more diverse set of features relative to the other conditions (TG and homogenous). In the TDG color condition, for example, half the targets transformed to red and half transformed green, whereas in the TG and homogenous trials, targets transformed to either all red or all green. Tracking objects that have more diverse features or that undergo more diverse feature transformations may be intrinsically harder. To address this confound, we conducted a third experiment that included a diversity condition, which was the same as the TDG condition, except that each distractor pair was drawn with a feature that differed from those of the targets. If low TDG performance in Experiments 1 and 2 resulted from featurally heterogeneous targets or more distracting feature transformations, then we would expect identical performance in the diversity and TDG conditions. On the other hand, if the difficulty of the TDG condition owed to automatic grouping between targets and distractors, then the diversity condition should yield higher accuracy than the TDG, as only the former incorporated deleterious grouping effects.

In this last experiment, we removed the contrast polarity, orientation, and depth conditions, as they failed the automaticity test of Experiment 2. We also excluded interpolation, because it had already been tested with a diversity condition in Keane et al. (2011, Experiments 1 and 4) and because we aimed to increase the number of trials per condition to more sensitively test for tracking differences. This left us with four feature types: color, shape, size, and their combination.

Method

Subjects.   Twenty-three UCLA undergraduates who did not take part in Experiments 1 and 2 participated for course credit. All had normal or corrected-to-normal vision and were subject to the same stereoacuity requirement as in the previous experiments.

Equipment and stimuli.   The equipment was the same as in the previous experiments. On diversity trials, the targets were the same as in TDG, but the distractor feature tokens were changed so that they did not pair up with targets (see Figure 1 and Table 1). For example, in the shape feature type, targets were triangles and circles, whereas distractors were diamonds and ovals. Objects that had the same features (e.g., the two circle targets) always appeared in opposite quadrants (e.g., top left and bottom right). For color, targets were red and green and distractors were blue and purple.

Table 1
Features of Targets and Distractors in Experiment 3

	Targets	Distractors
Color	Red, green	Blue, purple
Size	Large, small	Medium-small, medium-large
Shape	Triangle, circle	Diamond, oval
Combination	Red triangle and green circle, or red circle and green triangle	Blue diamond and purple oval, or blue oval and purple diamond

For size, targets were large and small, whereas distractors were two different intermediate sizes. In the combination condition, two targets were triangles and two were squares; one of each shape was red and the other was green. The distractors in the combination condition consisted of two diamonds and two ovals, and one of each shape type was purple and one was blue. For example, the eight objects could be a red triangle, a red square, a green triangle, a green square (targets), and a purple diamond, a purple oval, a blue diamond, a blue oval (distractors).

Procedure.   The procedure was identical to the previous experiment except that half as many feature types were tested and the number of trials was doubled for each condition within a block.

Results and Discussion

Results are shown in Figure 5. One-tailed *t* tests were performed on each kind of feature. (Tests were one-tailed because prior research suggests that the diversity condition should be at least as good as a condition that involves no grouping and better than a TDG condition; see Horowitz et al., 2007, Experiment 5). The following were significant: color, $t(22) = 5.950$, $p < .0001$, Cohen's *d*: 0.905; shape, $t(22) = 2.677$, $p < .01$, Cohen's *d*: 0.320; and combination, $t(22) = 3.923$, $p < .001$, Cohen's *d*: 0.668. The effect of size was marginally significant, $t(22) = 1.567$, $p = .0657$, Cohen's *d*: 0.208.

The null effect for size should be considered cautiously because there may have been weak target-distractor grouping in the diversity condition. Specifically, two of the targets and two of the distractors were larger than average, while the remaining stimuli were smaller than average, and the targets and distractors may have paired up on that basis. For the remaining feature types (color, shape, and combination), Experiment 3 showed that poorer performance in the TDG cannot be blamed on the heterogeneity of target features; it instead must owe to automatic feature-based grouping. A related implication is that a variety of features are automatically encoded for many objects at a time, even when those objects and features are task-irrelevant.

A caveat: There is some evidence that interference from the TDG condition decreased over the course of the experiment, suggesting that subjects were learning to strategically modulate the effect of target features, even if they could not ignore the feature information entirely. Accordingly, questionnaire data indicated that subjects were aware that the targets always shared similar features that were distinct from the distractors and that targets always appeared on opposite corners of the display. In order to assess whether grouping effects diminished over the course of the experiment, we plotted the accuracy difference between the diversity and TDG conditions as a function of trial number (see Figure 6). Specifically, for each subject and feature type, we subtracted the number correct for the *i*th diversity trial and the *i*th TDG trial for each of the 14 pairs of trials. These differences were then averaged across subjects and the resulting series was fit with a linear regression for each feature type (see Table 2). The slope of the fitted line was significantly different from zero only in the size condition ($p < .05$). We also fit the data with a multiple regression, using trial order (1 to 14) and feature types as factors. No interactions were significant, so we removed these and refit the model. The final model is shown in Table 3. These data suggest that
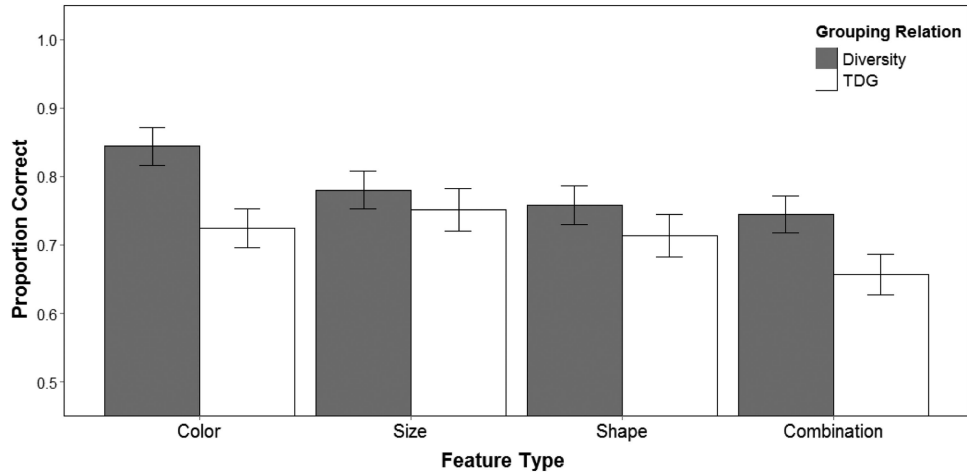
Figure 5. Proportion of correctly tracked targets from Experiment 3. The gray bars indicate performance in the diversity condition (targets and distractors do not share groupable features); the white bars show performance in the TDG condition, identical to Experiments 1 and 2. Grouping impaired performance for color, shape, and combination, and marginally for size.

learning and strategy reduce automatic grouping effects in some cases.

General Discussion

Contour interpolation is known to automatically group targets and distractors during multiple object tracking (Keane et al., 2011). We hypothesized a segmentation-relevance principle, according to

Figure 6. In Experiment 3, the average difference of total number of correctly identified targets between the diversity and TDG conditions as a function of trial number. A line was fit to the difference values for each feature type. The negative slopes of the lines indicate that—over time—subjects may be able to reduce the relative tracking disadvantage in the TDG conditions. However, only the slope for size was significantly different from zero (slope = −0.04, $p < .05$).

which features that segment the scene into discrete objects will be those that automatically influence tracking. Interpolation was specifically hypothesized to be unique among feature types in this capacity, because (a) it is phylogenetically, ontogenetically, and physiologically early; (b) it parses the scene into (trackable) units; and (c) it fills-in (rather than merely groups) spatially segregated objects. Over the course of three experiments, we replicated the finding that contour interpolation automatically groups targets and distractors; however, contrary to our hypothesis, we found that common shape, common color, and perhaps common size serve the same function. That is, when two targets and two distractors shared one feature token and the remaining four objects shared another, performance was inferior relative to when all targets shared a feature and all distractors shared another, when all objects had the same feature, and when targets and distractors had a variety of (nongrouping) features. These results—aside from revealing that features engender automatic grouping—suggest that a
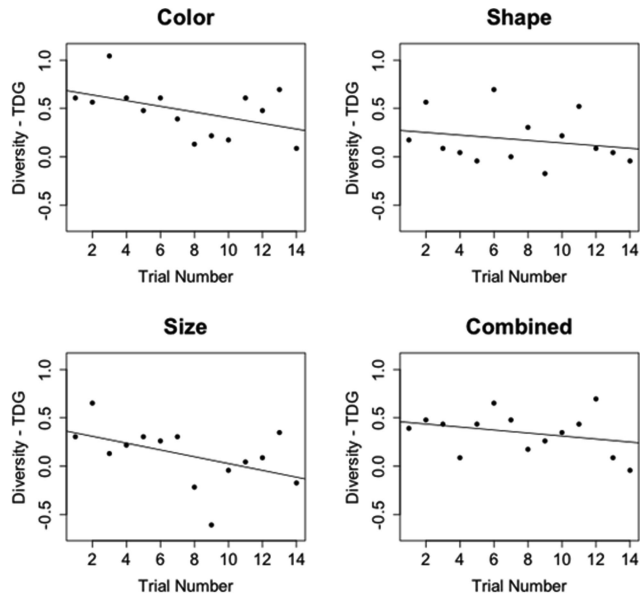
Table 2

Fit To TDG-Diversity Differences in Experiment 3

Condition	B	$SE\ B$	β
Color			
Intercept	0.70***	0.14	
Slope	−0.03	0.02	−0.10
Size			
Intercept	0.38**	0.13	
Slope	−0.04*	0.02	−0.12
Shape			
Intercept	0.28	0.15	
Slope	−0.01	0.02	−0.04
Combination			
Intercept	0.47**	0.15	
Slope	−0.02	0.02	−0.05

Note. Slopes and intercepts from fitting individual functions to the data from Experiment 3. Only the slope for the size condition was significantly different from 0 (uncorrected).
* $p < .05$. ** $p < .01$. *** $p < .001$. (all two-tailed)

Table 3
Simultaneous Fit to All Differences in Experiment 3

Condition	B	$SE\ B$	β
Constant	-0.46**	0.07	
Trial number	0.02**	0.01	0.08
Feature D1	-1.98**	0.06	-0.16
Feature D2	0.17*	0.06	0.13
Feature D3	0.10	0.06	0.08

Note. Results from multiple regression for Experiment 3.
* $p < .01$. ** $p < .001$ (all two-tailed).

variety of features are automatically encoded during tracking and that attentional allocation can be used as an index for perceptual organization.

Categorizing Features by Their Effects on Tracking

The feature types can be stratified into four categories. First, there are those features that are irrelevant to tracking. Strikingly, only one feature type in our study fell into this category—common orientation—but even this effect was marginal. It is possible that if the orientation differences were made more salient (either by increasing the orientation difference between tokens, or elongating the ovals), we might have observed a stronger effect. It should also be noted that orientation is not always irrelevant to MOT. In Experiment 4 of Keane et al. (2011), when two targets and two distractors were the vertices of two illusory quadrilaterals (as in the interpolation condition in Experiment 1 and 2 of this study), harmful grouping effects significantly weakened, as the Pac-Men were rotated up to 48 degrees from their initial (interpolating) positions. Therefore, orientation may become relevant to tracking primarily when it decides the trackable units that a scene contains.

The next category of feature types might be called *voluntary*, in that it benefits tracking (Experiment 1) by binding targets but it fails to impose deleterious target-distractor grouping effects (Experiment 2). This category includes stereoscopic depth and contrast polarity.

The third category—automatic nongrouping features—applies to the feature types that could aid tracking (Experiment 1), could not be ignored when they worked against the subject in Experiment 2, but did not impair performance in Experiment 3. Thus, these features probably harmed tracking in Experiment 2 because they broke up the unity of the target set but could not group targets and distractors together. By definition, this category includes only one of the tested feature types: size. This classification should be taken as provisional, given that the effect of size was marginal in Experiment 3, and given that there could have been some small degree of grouping in the diversity condition, as previously noted.

The final and most relevant category might be termed the *automatic grouping* feature types. As the name suggests, these are the feature types that produced an advantage in Experiment 1 and could not be ignored in either Experiment 2 or Experiment 3. This category includes—in addition to interpolation[1]—color, shape, combination, and possibly size.

Relation to Previous Results

Our results are seemingly inconsistent with several prior findings. Viswanathan and Mingolla (2002) found that performance *improved* when targets and distractors were each split between two depth planes (akin to our TDG condition) relative to when all objects shared the same depth plane (akin to our homogenous condition), whereas we found no difference between these two cases. Additionally, several previous studies found that when pairs of targets and distractors shared features (similar to our TDG condition), performance was the *same* relative to a homogenous condition in which all objects were identical (Horowitz et al., 2007; Makovski & Jiang, 2009a, 2009b; Viswanathan & Mingolla, 2002).

Among the variety of methodological differences between our studies and previous work, we focus on three that we believe to have the most explanatory power. First, in our studies, distinctive features were only available from the onset of motion to about 1 s before the objects stopped. In contrast, in all of the other studies, distinctive features were available at the target designation, allowing subjects to potentially utilize strategies to either mitigate or exploit the grouping, depending on the case. Second, previous studies employed at least twice as many trials/conditions (and often fewer subjects) in these studies than in ours (50 in Horowitz et al., 2007; 25 in Makovski & Jiang, 2009a; 15 in Makovski & Jiang, 2009b; 32 in Viswanathan & Mingolla, 2002; vs. 7 in our Experiment 2). The regression analyses in Experiment 3 (which had 14 trials per condition) suggest that subjects may notice the tendency for target-distractor grouping and learn to avoid such effects for some sorts of features. Finally, our objects were comparatively large (up to 3.7 degrees in diameter). Featural differences may have been considerably more salient—and hence harder to ignore—in our experimental setup.

The Usage of Object Features During Tracking

Previous research has shown that surface features tend to be ignored for many object continuity tasks (Feldman & Tremoulet, 2006; Flombaum et al., 2009; Mitroff & Alvarez, 2007; Pylyshyn, 2004; Scholl, 2007, 2009), especially when the task instructions do not explicitly ask the subject to attend to or use features. Some of our data support this view. Across all three experiments, objects in the combination condition dramatically and simultaneously changed in color, size, and shape twice in each trial, and subjects were still able to track at a level well above chance. Therefore, spatiotemporal continuity is critical for allowing the visual system to correspond multiple time slices of an entity over time, whereas featural identity appears to be less important (Bahrami, 2003; Feldman & Tremoulet, 2006). At the same time, featural effects have also been found in apparent motion, saccade, and occlusion studies (e.g., Moore, Mordkoff, & Enns, 2007). Hollingworth and Franconeri (2009) found that surface features can influence the perception of object correspondence, at least for brief occlusion events. If features matter for correspondence during occlusion, they very likely are important in other situations that require establishing continuity over time (e.g., continuous tracking).

---

[1] Interpolation produced a modest advantage in prior studies (~2%) that became significant only when data were averaged over 48 subjects. The present study was almost certainly underpowered to find this difference. It was argued that the small degree of facilitation owed to the deleterious effects of interdistractor interpolation (Keane et al., 2011, p. 695).

The present experiments show that surface features play an important role in unitizing a dynamic visual scene. In order for targets and distractors to group in our experiments, features must have been picked up "on the fly" after the targets were designated. Features derived not only from targets but also from distractors. Indeed, grouping could be switched on or off merely by modulating the appearance of the distractors, which is all the more surprising, given that distractors are plausibly inhibited during the process (Bettencourt, Michalka, & Somers, 2011; Bettencourt & Somers, 2009; Flombaum, Scholl, & Pylyshyn, 2008; Pylyshyn, 2006). The foregoing implies that if an object is whatever can be tracked and followed, then objecthood depends on featural identity. An alternative interpretation is that MOT is not about tracking *objects* in the ordinary sense of the term, but about tracking *collections* of related objects, sometimes referred to as *ensembles* (Halberda, Sires, & Feigenson, 2006). If so, features may be automatically employed during tracking under two circumstances: when they help specify the location, continuity, or extension of an object (as with interpolation), or when those features allow for the creation of ensembles, which can be composed of multiple objects. Automatic feature-based grouping may constitute an important means by which ensembles are composed.

The automatic grouping features do not appear to be additive. For example, in Experiment 1, the TG–TDG difference was about 21.2% for color and 21.1% for combination (which involved color, size, and shape). In Experiment 2, the TG–homogenous difference was 15.5% for color and 11.8% for combination. In both experiments, shape and size each had smaller, but still significant, effects. The lack of additivity cannot be ascribed to floor/ceiling effects because overall performance was close to 75% in the combination condition in both experiments, and because the differences in this condition never exceeded approximately 20%. We speculate that the visual system automatically uses only most the salient feature for grouping in these tasks, but of course this will need to be verified in future studies.

In the questionnaires, across all three experiments, very few subjects mentioned using an explicit grouping strategy. One subject in Experiment 1, two in Experiment 2, and six (out of 23) in Experiment 3 referred to object features as a part of a strategy to track objects. This makes it unlikely that subjects were using an explicit strategy to use features to group and track the objects. Removing these subjects and redoing the data analysis did not change any of the results. Furthermore, even if it could somehow be shown that subjects were unwittingly applying a nonoptimal cognitive strategy, that by itself would not rule out automaticity since other effects have been shown to be both automatic and cognitive in nature (Kahneman & Frederick, 2002).

Future Directions

Several future investigations follow naturally from the present. One question concerns whether learning and strategy dampen the default tendency to group. Experiment 3 suggests that strategy may be used to some extent, but it is unclear how this happens or under what circumstances. Future studies will need to examine the sorts of feature-based groupings that can be modulated by top-down strategy, the extent to which this modulation occurs, and the mechanisms and circumstances responsible.

Another question concerns the role of surfaces and contours in the interpolation condition: Do the observed effects of grouping owe to the surfaces (and attention spreading over them) or to the interpolation of contours across Pac-Men? Others have distinguished surface and contour mechanisms in vision (Breitmeyer & Tapia, 2011; Chen et al., 2007; Rogers-Ramachandran & Ramachandran, 1998; Yin, Kellman, & Shipley, 2000), and so it is at least possible that the former plays the most prominent role in automatic grouping.

Except for the interpolation condition, grouping in our study depended on a basic property: similarity. For example, color caused detrimental grouping only to the extent that the visual system represented red objects as similar to one another and as different from green. Therefore, tracking provides a means to address questions like the following: Are objects that share a shape but not a color more similar than those that share a color but not a shape (Goldstone & Son, 2005)? How different can two shades of color be and still be deemed perceptually similar? Is semantic similarity relevant to automatic grouping? MOT offers a new method to shed light on these and other questions by measuring the attentional "signatures" or correlates of perceptual grouping.

References

Bahrami, B. (2003). Object property encoding and change blindness in multiple object tracking. *Visual Cognition, 10,* 949–963. doi:10.1080/13506280344000158

Beck, J. (1966). Effect of orientation and of shape similarity on perceptual grouping. *Perception & Psychophysics, 1,* 300–302.

Bettencourt, K. C., Michalka, S. W., & Somers, D. C. (2011). Shared filtering processes link attentional and visual short-term memory capacity limits. *Journal of Vision, 11,* 22. doi:10.1167/11.10.22

Bettencourt, K. C., & Somers, D. C. (2009). Effects of target enhancement and distractor suppression on multiple object tracking capacity. *Journal of Vision, 9,* 9. doi:10.1167/9.7.9

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436. doi:10.1163/156856897X00357

Breitmeyer, B. G., & Tapia, E. (2011). Roles of contour and surface processing in microgenesis of object perception and visual consciousness. *Advances in Cognitive Psychology, 7,* 68–81. doi:10.2478/v10053-008-0088-y

Chan, L. K., & Hayward, W. G. (2009). Feature integration theory revisited: Dissociating feature detection and attentional guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 119–132. doi:10.1037/0096-1523.35.1.119

Chen, C. M., Lakatos, P., Shah, A. S., Mehta, A. D., Givre, S. J., Javitt, D. C., & Schroeder, C. E. (2007). Functional anatomy and interaction of fast and slow visual pathways in macaque monkeys. *Cerebral Cortex, 17,* 1561–1569. doi:10.1093/cercor/bhl067

Cohen, M. A., Pinto, Y., Howe, P. D., & Horowitz, T. S. (2011). The what-where trade-off in multiple-identity tracking. *Attention, Perception, & Psychophysics, 73,* 1422–1434.

Davis, G., & Driver, J. (1998). Kanizsa subjective figures can act as occluding surfaces at parallel stages of visual search. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 169–184. doi:10.1037/0096-1523.24.1.169

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113,* 501–517. doi:10.1037/0096-3445.113.4.501

Eagleman, D. M., & Sejnowski, T. J. (2007). Motion signals bias localization judgments: A unified explanation for the flash-lag, flash-drag, flash-jump, and Frohlich illusions. *Journal of Vision, 7,* 3. doi:10.1167/7.4.3

Earle, D. C. (1999). Glass patterns: Grouping by contrast similarity. *Perception, 28,* 1373–1382. doi:10.1068/p2986

Feldman, J., & Tremoulet, P. D. (2006). Individuation of visual objects over time. *Cognition, 99,* 131–165. doi:10.1016/j.cognition.2004.12.008

Flombaum, J. I., Scholl, B. J., & Pylyshyn, Z. W. (2008). Attentional resources in visual tracking through occlusion: The high-beams effect. *Cognition, 107,* 904–931. doi:10.1016/j.cognition.2007.12.015

Flombaum, J. I., Scholl, B. J., & Santos, L. R. (2009). Spatiotemporal priority as a fundamental principle of object persistence. In B. Hood & L. Santos (Eds.), *The Origins of Object Knowledge* (pp. 135–164). Oxford, UK: Oxford University Press.

Franconeri, S. L., Jonathan, S. V., & Scimeca, J. M. (2010). Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychological Science, 21,* 920–925. doi:10.1177/0956797610373935

Gilchrist, I. D., Humphreys, G. W., Riddoch, M. J., & Neumann, H. (1997). Luminance and edge information in grouping: A study using visual search. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 464–480. doi:10.1037/0096-1523.23.2.464

Gold, J. M., Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2000). Deriving behavioural receptive fields for visually completed contours. *Current Biology, 10,* 663–666. doi:10.1016/S0960-9822(00)00523-6

Goldstone, R. L., & Son, J. Y. (2005). Similarity. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 13–36). New York, NY: Cambridge University Press.

Gori, S., & Spillmann, L. (2010). Detection vs. grouping thresholds for elements differing in spacing, size and luminance. An alternative approach towards the psychophysics of Gestalten. *Vision Research, 50,* 1194–1202. doi:10.1016/j.visres.2010.03.022

Hadad, B. S., & Kimchi, R. (2008). Time course of grouping of shape by perceptual closure: Effects of spatial proximity and collinearity. *Perception & Psychophysics, 70,* 818–827. doi:10.3758/PP.70.5.818

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science: A Journal of the American Psychological Society, 17,* 572–576.

He, Z. J., & Nakayama, K. (1992). Surfaces versus features in visual search. *Nature, 359,* 231–233. doi:10.1038/359231a0

Hollingworth, A., & Franconeri, S. L. (2009). Object correspondence across brief occlusion is established on the basis of both spatiotemporal and surface features cues. *Cognition, 113,* 150–166. doi:10.1016/j.cognition.2009.08.004

Horowitz, T. S., Klieger, S. B., Fencsik, D. E., Yang, K. K., Alvarez, G. A., & Wolfe, J. M. (2007). Tracking unique objects. *Perception & Psychophysics, 69,* 172–184. doi:10.3758/BF03193740

Howard, C. J., & Holcombe, A. O. (2008). Tracking the changing features of multiple objects: Progressively poorer perceptual precision and progressively greater perceptual lag. *Vision Research, 48,* 1164–1180. doi:10.1016/j.visres.2008.01.023

Howe, P. D., Cohen, M. A., Pinto, Y., & Horowitz, T. S. (2010). Distinguishing between parallel and serial accounts of multiple object tracking. *Journal of Vision, 10,* 11. doi:10.1167/10.8.11

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge: Cambridge University Press.

Keane, B. P., Mettler, E., Tsoi, V., & Kellman, P. J. (2011). Attentional signatures of perception: Multiple object tracking reveals the automaticity of contour interpolation. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 685–698. doi:10.1037/a0020674

Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology, 23,* 141–221. doi:10.1016/0010-0285(91)90009-D

Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology, 15,* 483–524. doi:10.1016/0010-0285(83)90017-8

Makovski, T., & Jiang, Y. V. (2009a). Feature binding in attentive tracking of distinct objects. *Visual Cognition, 17,* 180–194. doi:10.1080/13506280802211334

Makovski, T., & Jiang, Y. V. (2009b). The role of visual working memory in attentive tracking of unique objects. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 1687–1697. doi:10.1037/a0016453

Mettler, E., Keane, B., & Kellman, P. (2008). Contour interpolation affects multiple object tracking. *Journal of Vision, 8,* 507. doi:10.1167/8.6.507

Mitroff, S. R., & Alvarez, G. A. (2007). Space and time, not surface features, guide object persistence. *Psychonomic Bulletin & Review, 14,* 1199–1204. doi:10.3758/BF03193113

Moore, C. M., Mordkoff, J. T., & Enns, J. T. (2007). The path of least persistence: Object status mediates visual updating. *Vision Research, 47,* 1624–1630. doi:10.1016/j.visres.2007.01.030

Moore, C. M., Yantis, S., & Vaughan, B. (1998). Object-based visual selection: Evidence from perceptual completion. *Psychological Science, 9,* 104–110. doi:10.1111/1467-9280.00019

Nakayama, K., Shimojo, S., & Silverman, G. H. (1989). Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception, 18,* 55–68. doi:10.1068/p180055

Nieder, A. (2002). Seeing more than meets the eye: Processing of illusory contours in animals. *Journal of Comparative Physiology. A, Neuroethology, Sensory, Neural, and Behavioral Physiology, 188,* 249–260. doi:10.1007/s00359-002-0306-x

Nothdurft, H. C. (1985). Orientation sensitivity and texture segmentation in patterns with different line orientation. *Vision Research, 25,* 551–560. doi:10.1016/0042-6989(85)90159-2

Oksama, L., & Hyönä, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach. *Visual Cognition, 11,* 631–671. doi:10.1080/13506280344000473

Oksama, L., & Hyönä, J. (2008). Dynamic binding of identity and location information: A serial model of multiple identity tracking. *Cognitive Psychology, 56,* 237–283. doi:10.1016/j.cogpsych.2007.03.001

Palmer, S., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review, 1,* 29–55. doi:10.3758/BF03200760

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. doi:10.1163/156856897X00366

Pessoa, L., Beck, J., & Mingolla, E. (1996). Perceived texture segregation in chromatic element-arrangement patterns: High intensity interference. *Vision Research, 36,* 1745–1760. doi:10.1016/0042-6989(95)00248-0

Pinto, Y., Howe, P. D., Cohen, M. A., & Horowitz, T. S. (2010). The more often you see an object, the easier it becomes to track it. *Journal of Vision, 10,* 4. doi:10.1167/10.10.4

Pylyshyn, Z. W. (2004). Some puzzling findings in multiple object tracking (MOT): I. Tracking without keeping track of object identities. *Visual Cognition, 11,* 801–822. doi:10.1080/13506280344000518

Pylyshyn, Z. W. (2006). Some puzzling findings in multiple object tracking (MOT): II. Inhibition of moving nontargets. *Visual Cognition, 14,* 175–198. doi:10.1080/13506280544000200

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3,* 179–197. doi:10.1163/156856888X00122

Rensink, R. A., & Enns, J. T. (1998). Early completion of occluded objects. *Vision Research, 38,* 2489–2505. doi:10.1016/S0042-6989(98)00051-0

Rock, I., Nijhawan, R., Palmer, S., & Tudor, L. (1992). Grouping based on phenomenal similarity of achromatic color. *Perception, 21,* 779–789. doi:10.1068/p210779

Rogers-Ramachandran, D. C., & Ramachandran, V. S. (1998). Psychophysical evidence for boundary and surface systems in human vision. *Vision Research, 38,* 71–77. doi:10.1016/S0042-6989(97)00131-4

Scholl, B. J. (2007). Object persistence in philosophy and psychology. *Mind & Language, 22,* 563–591.

Scholl, B. J. (2009). What have we learned about attention from multiple object tracking (and vice versa)? In D. Dedrick & L. Trick (Eds.), *Computation, cognition, and pylyshyn* (pp. 49–77). Cambridge, MA: MIT Press.

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition, 80,* 159–177. doi:10.1016/S0010-0277(00)00157-8

Sekuler, A. B., & Bennett, P. J. (2001). Generalized common fate: Grouping by common luminance changes. *Psychological Science, 12,* 437–444. doi:10.1111/1467-9280.00382

Shipley, T. F., & Kellman, P. J. (1992). Strength of visual interpolation depends on the ratio of physically specified to total edge length. *Perception & Psychophysics, 52,* 97–106. doi:10.3758/BF03206762

Störmer, V. S., Li, S.-C., Heekeren, H. R., & Lindenberger, U. (2011). Feature-based interference from unattended visual field during attentional tracking in younger and older adults. *Journal of Vision, 11,* 1–12. doi:10.1167/11.2.1

Suganuma, M., & Yokosawa, K. (2006). Grouping and trajectory storage in multiple object tracking: Impairments due to common item motions. *Perception, 35,* 483–495. doi:10.1068/p5487

Valenza, E., & Bulf, H. (2011). Early development of object unity: Evidence for perceptual completion in newborns. *Developmental Science, 14,* 799–808. doi:10.1111/j.1467-7687.2010.01026.x

Vickery, T. J., & Jiang, Y. V. (2009). Associative grouping: Perceptual grouping of shapes by association. *Attention, Perception, & Psychophysics, 71,* 896–909. doi:10.3758/APP.71.4.896

Viswanathan, L., & Mingolla, E. (2002). Dynamics of attention in depth: Evidence from multi-element tracking. *Perception, 31,* 1415–1437. doi:10.1068/p3432

von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science, 224,* 1260–1262. doi:10.1126/science.6539501

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5,* 495–501. doi:10.1038/nrn1411

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology, 24,* 295–340. doi:10.1016/0010-0285(92)90010-Y

Yin, C., Kellman, P. J., & Shipley, T. F. (2000). Surface integration influences depth discrimination. *Vision Research, 40,* 1969–1978. doi:10.1016/S0042-6989(00)00047-X