Post-print: Motion-Based Grouping Effect in Multi-Object Tracking

Authors: Hu Luming, Su Jing, Wei Liuqing, Zhang Xuemin

Date: 2018-09-07T00:00:00+00:00

Abstract

This study aims to investigate the influence of motion information on multiple object tracking, specifically examining whether a grouping effect based on motion information exists in multiple object tracking, and how different types of surface features (color and shape) affect this effect. Experiment 1 found that a grouping effect based on motion information exists in multiple object tracking, and that this grouping effect is automatic. Experiments 2 and 3 examined the stability of the grouping effect based on motion information under conditions of different color and shape combinations, respectively; both results indicated that the combination of different surface features does not interfere with the formation of the grouping effect based on motion information, but its effect size is reduced by the presence of surface features. Overall, an automatic grouping effect based on motion information exists in multiple object tracking, and this effect represents a concrete manifestation of the law of common fate in Gestalt principles, which can stably exist across different combinations of surface features.

Full Text

The Grouping Effect Based on Motion Information in Multiple Object Tracking

HU Luming[1], SU Jing[1], WEI Liuqing[1], ZHANG Xuemin[1,2]

[1]Beijing Key Laboratory of Applied Experimental Psychology; National Demonstration Center for Experimental Psychology Education (Beijing Normal University); Faculty of Psychology, Beijing Normal University

[2]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China

Abstract

This study investigated the influence of motion information on multiple object tracking, specifically examining whether a grouping effect based on motion information exists in multiple object tracking and how different types of surface features (color and shape) affect this effect. Experiment 1 found that a grouping effect based on motion information exists in multiple object tracking, and this grouping effect is automatic. Experiments 2 and 3 examined the stability of the motion-based grouping effect under different color and shape combination conditions, respectively. The results showed that different surface feature combinations did not interfere with the formation of the motion-based grouping effect, but the effect size was reduced by the presence of surface features. Overall, an automatic grouping effect based on motion information exists in multiple object tracking, and this effect represents a concrete manifestation of the Gestalt law of common fate, remaining stable across different surface feature combinations.

Keywords: multiple object tracking; common fate; motion information; grouping effect

1. Introduction

In daily life, people often need to use dynamic vision to track and observe the surrounding world. For instance, beach lifeguards must constantly monitor the status of people on the beach, and soccer players need to continuously track the positions of players and the ball on the field. Dynamic visual tracking ability is more complex and ecologically valid than static attentional perception, facilitating better processing and understanding of the dynamic objective world. In the 1980s, Pylyshyn and Storm (1988) proposed the multiple object tracking (MOT) task paradigm to systematically explore this dynamic visual ability. Since then, the multiple object tracking task has become a typical task for reflecting human capacity to acquire objects in spatiotemporally changing scenes. After nearly 30 years of research, it has been found that ordinary people can generally track 4–5 targets simultaneously, corresponding to an equivalent number of index tags (Pylyshyn & Storm, 1988). However, this capacity is still influenced by numerous factors, such as the number of targets or distractors (Oksama & Hyönä, 2004; Zhang, Lu, & Wei, 2011), object speed (Feria, 2013), motion trajectory (Luu & Howe, 2015), reference frame for motion (Huff, Meyerhoff, Papenmeier, & Jahn, 2010), spatial separation (Holcombe, Chen, & Howe, 2014), spatial symmetry features (Wang, Zhang, Li, & Lyu, 2016), surface features (Makovski & Jiang, 2009a; Lei, Wei, Lyu, Zhang, & Yan, 2016), categorical features (Wei, Zhang, Chuang, & Zhen, 2016), and identity transformations (Lyu, Hu, Wei, Zhang, & Talhelm, 2015; Hu, Lyu, Zhang, & Wei, 2018).

1.1 The Role of Motion Information in Multiple Object Tracking

Although multiple object tracking performance is affected by many factors, motion information is more fundamental and critical compared to other features. This is evident in two aspects. First, in typical multiple object tracking tasks where all moving objects are identical (Pylyshyn, 2004; Scholl, Pylyshyn, & Feldman, 2001), observers cannot use surface features to assist tracking but can still maintain simultaneous tracking of multiple targets effectively. Second, even when surface features change constantly (Makovski & Jiang, 2009b; Hu et al., 2018) or are occluded (Franconeri, Pylyshyn, & Scholl, 2012; Scholl & Pylyshyn, 1999), observers cannot effectively use surface features to assist tracking but can still rely on spatiotemporal characteristics to establish object continuity and maintain simultaneous tracking of multiple moving objects. Therefore, the visual system' s dependence on motion information is relatively strong in multiple object tracking (Keane & Pylyshyn, 2006; Zhang, Yao, & Lu, 2008).

Motion information concerns how current objects move (Fencsik, Klieger, & Horowitz, 2007) and can include comprehensive information such as position, speed, and direction (Suganuma & Yokosawa, 2006). Early on, Yantis (1992) and Scholl and Pylyshyn (1999) found that when visual objects disappeared or were occluded, they could still be tracked as continuous existences, meaning that objecthood was maintained (Scholl & Pylyshyn, 1999; Yantis, 1992). Even when the disappearance time was extended to between 150-900 ms, tracking performance showed no difference (Keane & Pylyshyn, 2006). When investigating the reasons for target recovery, Horowitz, Birnkrant, Fencsik, Tran, and Wolfe (2006) speculated that this target recovery ability mainly triggered an offline storage mechanism when target stimuli suddenly disappeared from view, with the visual system storing target information at the moment of disappearance in this offline memory (Horowitz et al., 2006). When targets reappeared, observers could match the current display with the stored information to recover targets. Regarding the information stored by this hypothesized offline storage mechanism, Keane and Pylyshyn (2006) believed it mainly consisted of target position information, while Fencsik et al. (2007) further used experiments to demonstrate that in addition to target position information, it should also encompass target motion direction and trajectory information (Fencsik, Horowitz, Kliege, & Wolf, 2004; Fencsik et al., 2007).

In addition to assisting target recovery, motion information can also serve as an auxiliary cue to influence multiple object tracking performance. First, Horowitz and Cohen (2010) required observers to judge the motion direction of targets after the motion phase in a typical MOT task, and the results showed that observers could identify it accurately. Second, StClair, Huff, and Seiffert (2010) also used a typical MOT paradigm and found that in real-time continuous motion of visible moving objects, texture background motion information significantly affected multiple object tracking performance (StClair et al., 2010). Compared to a static texture background, texture background motion opposite to target motion direction hindered multiple object tracking performance, while texture background motion in the same direction as target motion facilitated it. This suggests that motion information from moving objects can be used to promote multiple object tracking performance. Third, when moving objects moved linearly and their positions were predictable, observers could use motion information to assist tracking to some extent. However, when moving objects changed direction randomly and their positions were unpredictable, observers could not utilize motion information (Howe & Holcombe, 2012a).

Motion information helps establish object continuity, but whether it can promote multiple object tracking performance depends on whether it can be effectively organized and recognized. When motion information is randomly distributed in multiple object motion without regularity, the utilization degree of motion information is low, as in typical multiple object tracking tasks. However, when motion information is highly organized or highly recognizable, the facilitating effect of motion information is relatively obvious, such as with predictable motion positions and multiple object motion without directional information conflict. Suganuma and Yokosawa (2006) conducted a study focusing on how the organization of motion information affects multiple object tracking performance. The study selected four different motion trajectories as experimental conditions (see Figure 1 [Figure 1: see original paper]), namely the chase condition, the parallel condition, the trail condition, and the baseline random condition. In the chase condition, a target and a distractor first separated by a certain distance, after which either of them moved first (the chased object or leading object), and the other subsequently moved in the same direction and speed as the chased object (the chasing object or led object). The motion direction of the two objects was the same as the direction of the line connecting their centers. The parallel condition was the same as the chase condition in the initial stage, with two objects first separated by a certain distance and then moving in the same direction sequentially, but the front and rear objects differed by at least 30° in the same direction orientation, meaning they moved parallel to each other but with non-overlapping paths. The trail condition was the same as the chase condition, but while the leading object' s direction and speed were fixed and unchanging in the chase condition, the leading object in the trail condition was allowed to change direction and speed randomly. The random condition was consistent with typical MOT tasks and served as the experimental baseline.

Through sequential comparisons, it was found that performance in the chase condition was significantly lower than in the baseline random condition, indicating that multiple object tracking performance was affected by object motion trajectory. Furthermore, both the parallel condition and the trail condition were significantly lower than the baseline random condition but showed no significant difference from the chase condition, respectively indicating that sharing motion direction and sharing motion trajectory both reduce multiple object tracking performance. Therefore, when targets and distractors share certain motion information, such as direction or trajectory, it will impair multiple object tracking performance, even if observers do not successfully detect this relationship (Suganuma & Yokosawa, 2006). This study also implies that even in the absence of other effective grouping cues, different types of common motion information seem to be able to form grouping as a holistic object representation. However, this study mainly used the grouping hypothesis to explain the effect produced by common motion information, suggesting that high-level perceptual organization can use common motion information to maintain tag-target updating, representing a process from phenomenon to theoretical explanation. It did not directly elaborate on motion information itself as a type of feature forming grouping under different circumstances and its impact on multiple object tracking performance. Specifically, whether motion information features can form grouping both within targets and between targets and distractors, similar to other surface features (such as color, shape), and whether these two types of grouping produce different effects on multiple object tracking performance. In comparison, the content explored in this study is more direct and specific than that of Suganuma and Yokosawa (2006).

[Figure 1: see original paper] Three types of motion trajectories (Suganuma & Yokosawa, 2006)

Note: a. Chase motion; b. Parallel motion; c. Trail motion. Dashed lines represent the motion trajectories of solid objects, while solid lines represent the motion trajectories of hollow objects. The bolded portions of dashed and solid lines indicate the critical phase of motion matching. The motion positions in the above images are captured from the first frame of the critical phase. Note that 8 or 10 identical objects appeared in the formal experiment, with no indicator lines present.

1.2 Grouping Effects in Multiple Object Tracking

Yantis (1992) first reported that grouping effects facilitate multiple object tracking performance, meaning that observers can top-down form virtual polygons at a high cognitive level to maintain multiple tracking targets. Specifically, in multiple object tracking, observers first automatically form grouping through stimulus-driven processes (group formation), and then maintain grouping through top-down goal-driven processes (group maintenance). Scholl et al. (2001) used target merging methods, connecting a target and a distractor with visible or invisible lines, and found that target merging technology caused the merged target and distractor to be automatically perceived as a holistic object, increasing tracking task difficulty. Subsequently, a series of multiple object tracking studies have also shown that moving objects in multiple object tracking can form perceptual grouping based on different characteristics, such as similarity (Feria, 2012; Howe & Holcombe, 2012b; Makovski & Jiang, 2009a, 2009b), interpolated contours (Keane, Mettler, Tsoi, & Kellman, 2011), stereo depth and contrast polarity (Erlikhman, Keane, Mettler, Horowitz, & Kellman, 2013), symmetry regularity (Wang et al., 2016), conceptual categories (Wei et al., 2016), and combined features (Erlikhman et al., 2013; Howe & Holcombe, 2012b).

Makovski and Jiang (2009a, 2009b) showed that when targets and distractors share color features, it impairs attentional tracking. Feria (2012) demonstrated that when distractors differ from targets in shape, color, or stationary state, the impairment to tracking performance is smaller than when distractors are identical to targets. Howe and Holcombe (2012b) further showed that when distractors do not share combined features with targets, combined features can also facilitate multiple object tracking performance. Keane et al. (2011) found that in addition to surface features, interpolated contour features can also bind targets and distractors, subsequently causing impaired tracking performance, and this binding is automatic. To verify whether this automatic grouping effect is unique or universally present, Erlikhman et al. (2013) more comprehensively examined grouping effects across eight features, including color, contrast, orientation, size, shape, stereo depth, interpolation, and feature combinations (color, shape, size). The results showed that most features forming inter-target grouping could significantly improve tracking performance, except for grouping based on orientation and interpolation features. However, when forming target-distractor grouping based on specific features, it significantly impaired tracking performance, including color, size, shape, feature combinations, and interpolation. Therefore, the vast majority of features can produce grouping effects based on feature similarity, even when these features are irrelevant or contrary to the current task, and in dynamic tracking, feature similarity-based grouping effects seem to favor automatic and parallel processing. Wang et al. (2016) recently also found that regular features, such as spatial symmetry, can affect multiple object tracking performance. When targets and targets are in symmetric positions, multiple object tracking performance improves, whereas when targets and distractors are in symmetric positions, performance decreases. Furthermore, symmetry-based grouping effects can be additive with color- or shape-based grouping effects, further improving multiple object tracking performance. This indicates that grouping effects between two different types of features can operate independently without interfering with each other.

While the above studies have discovered numerous grouping effects based on different surface features, whether basic motion and spatiotemporal feature attributes can also produce similar grouping effects remains to be further investigated in detail. Motion and spatiotemporal features play important roles in establishing object continuity, and theoretically, different types of motion grouping effects will influence the process of establishing object continuity. Therefore, this experiment hypothesizes that if motion information can assist observers in forming grouping within the target group, this grouping effect will be more conducive to establishing object continuity. Conversely, if motion information assists observers in forming grouping between targets and distractors, this grouping effect will more likely interfere with the establishment of object continuity. This can explore whether motion information-based grouping effects exist, and also examine whether motion information-based grouping effects tend toward automatic processing. In Experiment 1, three different motion information conditions were set up, corresponding to intra-target paired motion, target-distractor paired motion, and random motion (baseline condition), to investigate this grouping effect. According to Keane et al. (2011) and Erlikhman et al. (2013), for a feature to form automatic grouping during multiple object tracking, two conditions must be met. First, performance when grouping forms within targets must be better than performance in random motion state. Second, performance when grouping forms between targets and distractors must be worse than performance in random motion state. Meeting the first condition indicates that the feature can effectively form perceptual grouping, meaning observers can effectively distinguish targets from distractors based on this feature to improve tracking performance. However, meeting the first condition does not mean the feature is automatically processed; it could also be voluntarily processed. Therefore, the second condition must also be met. If the second condition is also satisfied, it indicates that this feature can not only effectively distinguish targets from distractors but also effectively interfere with the visual object independence of targets, hindering targets from being distinguished from distractors. Moreover, this effect cannot be changed by observers' voluntary inhibition but is a bottom-up automatic process that inevitably causes visual confusion. Therefore, in Experiment 1, if tracking performance in both the intra-target paired motion condition and the target-distractor paired motion condition is significantly different from random motion, it indicates that paired motion can affect multiple object tracking, meaning observers can use motion information to form grouping. Moreover, if tracking performance in the intra-target paired motion condition is significantly higher than in random motion and tracking performance in the target-distractor paired motion condition is significantly lower than in random motion, it indicates that this grouping effect also supports bottom-up automatic processing.

Additionally, considering that a large number of moving objects in daily life have their own surface feature information, the current study further investigated whether the motion information-based grouping effect can stably exist in multiple object tracking tasks with surface features, that is, whether surface features affect the formation of motion information-based grouping effects. Specifically, in Experiment 1, the surface feature information available to observers was consistent, meaning observers mainly relied on motion information to maintain multiple object tracking. However, when surface feature information appears, will the utilization degree of motion information be affected? Pylyshyn (2004) believed that surface features have a small effect on multiple object tracking, while Makovski and Jiang (2009b) and Feria (2012) believed that surface features play an important role in multiple object tracking. Therefore, examining motion information-based grouping effects under different surface features has important theoretical and practical significance. In Experiments 2 and 3, we explored the stability of motion information-based grouping effects by manipulating different color and shape combination features, including three combination conditions: identical feature condition, distinct feature condition, and mixed feature condition. If motion information-based grouping effects exist under different color and shape feature conditions, there will be sufficient evidence to support that motion information-based grouping effects stably exist in multiple object tracking. Overall, the current study is of considerable significance for understanding the role of motion information in multiple object tracking and for further understanding the impact of perceptual grouping effects on multiple object tracking.

2. Experiment 1: The Effect of Motion Information-Based Grouping on Multiple Object Tracking Performance

Motion information is widely used in multiple object tracking, such as in target recovery and assisted tracking, but its utilization degree varies depending on organizational forms. Although grouping effects have been proven to widely exist in numerous surface features, for basic spatiotemporal features, there is still no complete direct evidence indicating the existence of motion information-based grouping effects and whether this effect is driven by bottom-up automatic processing. The essence of automatic grouping effects in multiple object tracking is that intra-target grouping facilitates multiple object tracking performance, while target-distractor grouping impairs performance (Erlikhman et al., 2013; Keane et al., 2011; Wang et al., 2016). Therefore, this experiment will demonstrate the existence of motion information-based grouping effects and their automatic processing characteristics by examining differences in multiple object tracking performance between intra-target motion grouping, target-distractor motion grouping, and random motion conditions.

Sixteen participants were recruited from universities in the Beijing area, including 6 males and 10 females, with a mean age of 21.38 ± 2.17 years. All participants had normal or corrected-to-normal vision. They signed informed consent forms before the experiment and received corresponding compensation after the experiment.

The experimental display was a 17-inch CRT monitor with a screen resolution of 1024×768 pixels (1 pixel = 0.032 cm) and a vertical refresh rate of 85 Hz. The experimental program was written using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) based on Matlab 2012b (www.mathworks.com).

During the formal experiment, participants sat approximately 57 cm in front of the computer screen. The stimulus presentation area was a black [RGB (0, 0, 0)] square frame of 900×675 pixels (36°×27°) at the center of the screen. Both the background and motion area were filled with natural gray [RGB (128, 128, 128)]. The tracking objects were 12 black rings with a diameter of 30 pixels (8 pixels wide), including 6 targets and 6 distractors. At the beginning of the experiment, the initial positions of tracking objects were randomly assigned, with an average motion speed of 8.5°/s. The direction was not changed randomly during motion; objects rebounded after hitting the outer walls but did not rebound upon contact with each other.

The experimental design was a single-factor three-level within-subjects design (see Figure 2). The independent variable was the pairing form of tracking objects' motion trajectories. Condition 1 was target-target paired chase motion (target-target pairing condition), where one target moved in front while another target followed closely behind. Condition 2 was target-distractor paired chase motion (target-distractor pairing condition), where one target moved in front while a distractor followed closely behind. Condition 3 was the control condition (random no-pairing condition), where targets and distractors moved independently and randomly, consistent with typical MOT tasks. The dependent variable was tracking accuracy.

The chase motion began in a paired state, with the front and rear moving objects separated by three times the center-to-center distance (90 pixels, 3.6 degrees). Thereafter, the chasing target followed the motion trajectory of the chased target. The chasing and chased targets had consistent motion speed and direction, with their motion direction pointing toward the line connecting their centers. The front object rebounded when hitting the black border, and the rear chasing object rebounded at the same position along the front object' s trajectory, then continued chasing the front object.

Note: The red dashed circles indicate tracking targets. Dashed arrows represent paired motion relationships; in each paired motion relationship, the object pointed to by the arrow is the leader, and the object at the arrow' s starting point is the follower. The follower' s motion trajectory is consistent with the leader' s, showing a chasing state. In formal motion, red dashed circles and dashed arrows are not displayed.

[Figure 2: see original paper] Schematic diagram of three experimental conditions

At the beginning of the experiment, instructions were automatically presented on the screen. After confirming that participants understood the task, they pressed the spacebar to continue. Then, 12 rings with a diameter of 30 pixels were presented simultaneously on the screen, with 6 rings selected as targets. A red ring with a diameter of 40 pixels (4 pixels wide) appeared around the targets and flashed 4 times for 2000 ms. After flashing, the marking state was maintained for 1000 ms (i.e., the red ring continued to appear around tracking objects). After the marking phase, all tracking objects entered the motion phase. In paired chase motion, the center-to-center distance between front and rear paired objects was 90 pixels (three times the ring diameter). Paired rings moved in the same direction and speed (appearing as one object chasing another). The front object rebounded when hitting the black border, and the rear chasing object rebounded at the same position along the front object's trajectory, then continued following the front object. During motion, rings could occlude each other without collision rebound. In random motion without pairing, all rings moved randomly and independently, repelling each other at a center-to-center distance of three times the ring diameter. The occlusion settings differed between paired motion and random motion. Occlusion was allowed in the former setting to avoid large-scale collision rebound phenomena that would affect paired tracking effectiveness. The latter setting used collision rebound without occlusion because half of the moving objects in paired chase motion were in a non-contact state with a spacing of three times the object diameter. To counteract the potential gain effect caused by this phenomenon in paired motion, random motion was set to rebound before contact. This setting is more conservative and less likely to cause false-positive results. The entire motion phase lasted 10 s, entering the paired state directly from the start of motion. To prevent participants from relying only on chased object information and ignoring chasing object information in paired chase motion, all objects were unpaired 1 s before the end of paired chase motion and entered random motion. After the motion phase, participants were required to select all targets with the mouse. A red ring with a diameter of 40 pixels appeared around selected rings, and selections could be changed. After selecting all 6 targets, participants pressed the spacebar to proceed to the next trial.

Before the formal experiment, participants practiced for at least 4 trials with accuracy feedback until they clearly understood the experimental task. The three experimental conditions were presented in mixed random order, with 20 trials per condition, totaling 60 trials. The entire experiment lasted approximately 20 minutes. The experimental program automatically recorded response data.

2.5 Results and Analysis

Repeated measures ANOVA revealed significant differences between different motion patterns (see Figure 3 [Figure 3: see original paper]), $F(2, 30) = 22$, $p < 0.001$, $\text{p}^2 = 0.6$. Post-hoc pairwise comparisons (Bonferroni corrected) further showed that the target-target pairing condition was significantly higher than the target-distractor pairing condition, $t(15) = 7.51$, $p < 0.001$, Cohen' s $d = 1.88$, and the random no-pairing condition, $t(15) = 3.23$, $p = 0.006$, Cohen' s $d = 0.81$. Moreover, the target-distractor pairing condition was significantly lower than the random no-pairing condition, $t(15) = -3.06$, $p = 0.008$, Cohen' s $d = 0.76$.

[Figure 3: see original paper] Tracking accuracy for three motion conditions (error bars represent +1 SE)

Experiment 1 found that tracking accuracy was highest in the target-target pairing condition, followed by the random no-pairing condition, and lowest in the target-distractor pairing condition. That is, compared to the baseline condition, when distractors share motion information with targets, tracking performance is impaired, consistent with previous research (Makovski & Jiang, 2009b; Suganuma & Yokosawa, 2006). Compared to the baseline condition, when targets share motion information with each other, tracking performance can be improved (Howe & Holcombe, 2012b). This indicates that the grouping effect based on motion trajectory is also automatic (Erlikhman et al., 2013). This suggests that motion information, like surface features such as color, size, and shape, is also an important factor in forming grouping effects.

3. Experiment 2: The Stability of Motion Information-Based Grouping Effects Under Different Color Feature Combinations

Experiment 1 found that a grouping effect based on motion information exists in multiple object tracking, and this grouping effect is automatically processed. In Experiment 1, all moving objects had identical surface features, meaning they had the same color and shape. However, in the real world, moving objects have certain surface features. Does the motion information-based grouping effect stably exist under different surface features? Specifically, do different surface features affect the formation of motion information-based grouping effects? Based on this, Experiment 2 selected color features as surface features and divided surface feature combinations into three situations: same color between targets and distractors (identical color condition), different colors between targets and distractors (distinct color condition), and mixed colors between targets and distractors (mixed color condition). Under these three surface feature conditions, we examined whether the motion information-based grouping effect still stably existed.

Eighteen participants were recruited from universities in the Beijing area, including 7 males and 11 females, with a mean age of 21.25 ± 1.52 years. All participants had normal or corrected-to-normal vision. They signed informed consent forms before the experiment and received corresponding compensation after the experiment.

The experimental apparatus and program were the same as in Experiment 1. The experimental materials were 12 solid circles with a diameter of 30 pixels, including 6 targets and 6 distractors. The colors filled in the solid circles were as follows (see Figure 4 [Figure 4: see original paper]): green [RGB (0, 255, 0)], magenta [RGB (255, 0, 255)], cyan [RGB (0, 255, 255)], blue [RGB (0, 0, 255)], red [RGB (255, 0, 0)], orange [RGB (255, 165, 0)], and deep pink [RGB (255, 20, 147)]. Target and distractor colors were randomly selected from the above colors. The average motion speed of moving objects was 10.2°/s. Other parameters were the same as in Experiment 1.

[Figure 4: see original paper] Experimental color materials (see color version in electronic edition)

This experiment used a 2 (motion information grouping: target-target pairing condition vs. target-distractor pairing condition) × 3 (color grouping: identical color condition, distinct color condition, mixed color condition) within-subjects design. The experimental combination design is shown in Table 1 and Figure 5 [Figure 5: see original paper].

Table 1 Experimental Design Description for Experiment 2

Combination Condition	Explanation
Target-Target Pairing	Target-target paired chase motion, with targets and distractors in the same color
Target-Target Pairing	Target-target paired chase motion, with targets in one color and distractors in another color
Target-Target Pairing	Target-target paired chase motion, with half targets and half distractors in one color and the other half targets and distractors in another color
Target-Distractor Pairing	Distractor-target paired chase motion, with targets and distractors in the same color
Target-Distractor Pairing	Distractor-target paired chase motion, with targets in one color and distractors in another color
Target-Distractor Pairing	Distractor-target paired chase motion, with half targets and half distractors in one color and the other half targets and distractors in another color

[Figure 5: see original paper] Schematic diagram of experimental conditions (color + motion information, see color version in electronic edition. Circles within dashed square frames are marked as tracking targets, others are distractors. In the formal experiment, red rings were used for marking during the marking phase; this is only illustrative here. Dashed arrows indicate paired motion relationships; in each paired motion relationship, the object pointed to by the arrow is the leader, and the object at the arrow' s starting point is the follower. The follower' s motion trajectory is consistent with the leader' s, showing a chasing state. Dashed frames and dashed arrows are not displayed during formal motion)

The experimental procedure and parameters were basically the same as in Experiment 1. However, color was added as a surface feature in each trial, and at the end of motion, solid black circles of the same size (30 pixels) were used for masking to prevent participants from using memory strategies for selection. Participants first completed the target-target pairing condition under three color combinations, then completed the target-distractor pairing condition under three color combinations. In each chase condition, the three color combinations were randomly mixed. Each combination condition was repeated for 20 trials, totaling 120 trials. The entire experiment lasted approximately 40 minutes. The experimental program automatically recorded response data.

3.5 Results and Analysis

Tracking accuracy under different motion information and color combination conditions is shown in Figure 6 [Figure 6: see original paper]. Repeated measures ANOVA revealed a significant main effect of motion information grouping, $F(1, 17) = 28.76$, $p < 0.001$, $p^2 = 0.63$. This indicates that overall, tracking accuracy in the target-target pairing condition was significantly higher than in the target-distractor pairing condition. The main effect of color grouping was significant, $F(1.37, 23.37) = 213.86$, $p < 0.001$, $p^2 = 0.93$ (Greenhouse-Geisser corrected). This indicates that different color grouping treatments in the experiment were effective. The interaction between the two factors was also significant, $F(2, 34) = 9.44$, $p < 0.001$, $p^2 = 0.36$. This indicates that the differences in tracking accuracy between target-target pairing and target-distractor pairing conditions under different color feature groupings were not the same.

[Figure 6: see original paper] Tracking accuracy based on motion grouping under three color grouping conditions (error bars represent ±1 SE)

Further simple effect tests (Bonferroni corrected) under different color grouping conditions found that tracking accuracy in the target-target pairing condition was significantly higher than in the target-distractor pairing condition under the identical color condition ($t(17) = 6.65$, $p < 0.001$, Cohen' s $d = 1.57$), distinct color condition ($t(17) = 2.72$, $p = 0.015$, Cohen' s $d = 0.64$), and mixed color condition ($t(17) = 2.19$, $p = 0.043$, Cohen' s $d = 0.52$). This indicates that motion information-based grouping effects stably exist across different color feature groupings. On the other hand, simple effect tests (Bonferroni corrected) under different motion information grouping conditions found that the three color grouping conditions were significant in the target-target pairing condition, $F(2, 16) = 93.36$, $p < 0.001$, $p^2 = 0.92$. Specifically, the distinct color condition was significantly higher than the identical color condition ($t(17) = 12.39$, $p < 0.001$, Cohen's $d = 2.92$) and the mixed color condition ($t(17) = 14.08$, $p < 0.001$, Cohen' s $d = 3.32$). The mixed color condition was significantly lower than the identical color condition, $t(17) = -2.9$, $p = 0.03$, Cohen' s $d = 0.68$. The three color grouping conditions were also significant in the target-distractor pairing condition, $F(2, 16) = 173.36$, $p < 0.001$, $p^2 = 0.96$. Specifically, the distinct color condition was significantly higher than the identical color condition ($t(17) = 18.79$, $p < 0.001$, Cohen' s $d = 4.43$) and the mixed color condition ($t(17) = 13.13$, $p < 0.001$, Cohen' s $d = 3.1$). The mixed color condition showed no significant difference from the identical color condition, $t(17) = -2.15$, $p = 0.139$, Cohen' s $d = -0.51$. This indicates that in feature combination conditions, the distinct color condition can also significantly facilitate multiple object tracking performance, while the mixed color condition cannot facilitate and may even impair tracking performance. That is, to a certain extent, color features and motion information can be utilized simultaneously.

Furthermore, to analyze the effect of color features on motion information, the tracking performance in the target-distractor pairing condition was subtracted from the target-target pairing condition under the three color grouping conditions, yielding gain scores of 10.1%, 3.6%, and 3.4% for the three color grouping conditions, respectively. Using the gain score under the identical color condition as a baseline, paired-sample t-tests with the other two gain scores revealed that the gain scores in both the distinct color condition ($t(17) = -3.73$, $p = 0.002$, Cohen’ s $d = -0.88$) and the mixed color condition ($t(17) = -4.26$, $p = 0.001$, Cohen’ s $d = -1$) were significantly lower than in the identical color condition. However, there was no significant difference in gain scores between the distinct color and mixed color conditions, $t(17) = 0.1$, $p = 0.924$, Cohen’ s $d = 0.02$. This indicates that although motion information-based grouping effects stably exist across different color feature groupings, their effect sizes are significantly reduced by surface features. When color grouping is advantageous, such as in the distinct color condition, the reduction is due to color grouping gains. When color grouping is disadvantageous, such as in the mixed color condition, the reduction is due to color mixing interference.

In different color combination conditions, target-target pairing was always more conducive to tracking than target-distractor pairing, indicating that motion information-based grouping effects are stable. Color-based grouping effects did not interfere with the formation of motion information-based grouping effects. However, compared to when targets and distractors had identical colors, the presence of color features significantly reduced the difference between target-target pairing and target-distractor pairing conditions. Although the reasons for the reduction differed—gain reduction in the former and interference reduction in the latter—this demonstrates on one hand that motion information-based grouping effects have stability and can still function even when color-based grouping effects exist. On the other hand, it also shows that although color-based grouping effects did not interfere with the formation of motion information-based grouping effects, motion information-based grouping effects were still influenced by color features.

4. Experiment 3: The Stability of Motion Information-Based Grouping Effects Under Different Shape Feature Combinations

Experiment 2 examined motion information-based grouping effects under different color feature combinations and still found that intra-target grouping yielded better multiple object tracking performance than target-distractor grouping. However, color features are only one type of surface feature, and shape features are needed for further verification. Shape features, like color features, are external features widely present on object surfaces, but the cognitive processing of shape features differs from that of color features. Therefore, Experiment 3 further examined motion information-based grouping effects under different shape grouping conditions.

Forty participants were recruited from universities in the Beijing area (including 11 males and 29 females, with a mean age of 22.38 ± 2.89 years). All participants had normal or corrected-to-normal vision. They signed informed consent forms before the experiment and received corresponding compensation after the experiment.

The experimental apparatus and program platform were the same as in Experiment 2. However, the experimental materials were no longer fixed as solid circles but used circles, isosceles triangles, squares, and pentagons as stimulus materials according to shape features. The latter three were inscribed geometric figures of the circle (see Figure 7 [Figure 7: see original paper]), with the reference circle diameter remaining at 30 pixels. In the experiment, the shapes used for targets and distractors were randomly selected from the above four shapes. The average motion speed of moving objects in the experiment was 10.2°/s, consistent with Experiment 2.

[Figure 7: see original paper] Experimental shape materials (in formal experiments, the circumscribed circles are not displayed)

The experimental procedure and parameters were the same as in Experiment 2, but shape was changed as a surface feature in each trial. At the end of motion, circumscribed black solid circles of the same size (same size as the reference circle, i.e., 30 pixels in diameter) were used for masking to prevent participants from using memory strategies for selection. In each chase condition, the three shape combinations were randomly mixed. Each combination condition was repeated for 20 trials, totaling 120 trials. The entire experiment lasted approximately 40 minutes. The experimental program automatically recorded response data.

4.5 Results and Analysis

Tracking accuracy under different motion information and shape combination conditions is shown in Figure 9 [Figure 9: see original paper]. Repeated measures ANOVA revealed a significant main effect of motion information grouping, $F(1, 39) = 195.52$, $p < 0.001$, $\text{p}^2 = 0.83$. This indicates that overall, tracking accuracy in the target-target pairing condition was significantly higher than in the target-distractor pairing condition. The main effect of shape grouping was significant, $F(2, 78) = 241.57$, $p < 0.001$, $\text{p}^2 = 0.86$. This indicates that different shape grouping treatments in the experiment were effective. The interaction between the two factors was significant, $F(2, 78) = 5.23$, $p = 0.007$, $\text{p}^2 = 0.12$. This indicates that the differences in tracking accuracy between target-target pairing and target-distractor pairing conditions under different shape feature groupings were significantly different.

[Figure 9: see original paper] Tracking accuracy based on motion grouping under three shape grouping conditions (error bars represent ±1 SE)

Simple effect tests (Bonferroni corrected) under different shape feature grouping conditions found that tracking accuracy in the target-target pairing condition was significantly higher than in the target-distractor pairing condition under the identical shape condition ($t(39) = 13.91$, $p < 0.001$, Cohen' s $d = 2.20$), distinct shape condition ($t(39) = 10.67$, $p < 0.001$, Cohen' s $d = 1.69$), and mixed shape condition ($t(39) = 10.59$, $p < 0.001$, Cohen' s $d = 1.67$). This indicates that motion information-based grouping effects stably exist across different shape feature groupings. On the other hand, simple effect tests (Bonferroni corrected) under different motion information grouping conditions found that the three shape grouping conditions were significant in the target-target pairing condition, $F(2, 38) = 126.05$, $p < 0.001$, $p^2 = 0.87$. Specifically, the distinct shape condition was significantly higher than the identical shape condition ($t(39) = 13.58$, $p < 0.001$, Cohen' s $d = 2.15$) and the mixed shape condition ($t(39) = 15.99$, $p < 0.001$, Cohen' s $d = 2.53$). The mixed shape condition showed a marginally significant difference from the identical shape condition, $t(39) = -2.38$, $p = 0.067$, Cohen' s $d = -0.38$. The three shape grouping conditions were also significant in the target-distractor pairing condition, $F(2, 38) = 91.63$, $p < 0.001$, $p^2 = 0.83$. Specifically, the distinct shape condition was significantly higher than the identical shape condition ($t(39) = 13.30$, $p < 0.001$, Cohen' s $d = 2.10$) and the mixed shape condition ($t(39) = 12.93$, $p < 0.001$, Cohen' s $d = 2.05$). The mixed shape condition showed a significant difference from the identical shape condition, $t(39) = 2.56$, $p = 0.044$, Cohen' s $d = 0.4$. This indicates that in feature combination conditions, the distinct shape condition can also significantly facilitate multiple object tracking performance, while the mixed shape condition cannot facilitate and may even significantly impair tracking performance. That is, to a certain extent, shape features and motion information can also be utilized simultaneously.

Furthermore, subtracting the target-distractor pairing condition from the target-target pairing condition under the three shape grouping conditions yielded gain scores of 20.99%, 17.06%, and 16.64% for the three shape grouping conditions, respectively. Using the gain score under the identical shape condition as a baseline, paired-sample t-tests with the other two gain scores revealed that the gain scores in both the distinct shape condition ($t(39) = -2.41$, $p = 0.021$, Cohen' s $d = -0.38$) and the mixed shape condition ($t(39) = -3.98$, $p < 0.001$, Cohen's $d = -0.63$) were significantly lower than in the identical shape condition. However, there was no significant difference in gain scores between the distinct shape and mixed shape conditions, $t(39) = -0.25$, $p = 0.803$, Cohen' s $d = -0.04$. This indicates that although motion information-based grouping effects stably exist across different shape feature groupings, their effect sizes are significantly reduced by surface features. When shape grouping is advantageous, such as in the distinct shape condition, the reduction is due to shape grouping gains. When shape grouping is disadvantageous, such as in the mixed shape condition, the reduction is due to shape mixing interference.

In different shape combination conditions, target-target pairing was always more conducive to tracking than target-distractor pairing, again indicating that motion information-based grouping effects are stable. Shape-based grouping effects did not interfere with the formation of motion information-based grouping effects. However, compared to when targets and distractors had identical shapes, the presence of shape features also significantly reduced the difference between target-target pairing and target-distractor pairing conditions. Therefore, this demonstrates on one hand that motion information-based grouping effects have stability and can still function even when shape-based grouping effects exist. On the other hand, it also shows that although shape-based grouping effects did not interfere with the formation of motion information-based grouping effects, motion information-based grouping effects were still influenced by shape features. The results of Experiment 3 are consistent with those of Experiment 2, indicating that motion information-based grouping effects stably exist under different color or shape conditions, and the effect size is reduced due to the presence of surface features.

5. General Discussion

Experiment 1 demonstrated the existence of an automatic grouping effect based on motion information. Experiments 2 and 3 further examined the stability of this motion information-based grouping effect under different color and shape combination conditions. The results showed that different surface feature combinations did not interfere with the formation of motion information-based grouping effects, indicating that motion information-based grouping effects have stability, but their effect size is reduced by surface features.

5.1 Grouping Effects in Multiple Object Tracking and Gestalt Laws

The automatic grouping effect based on motion information found in the current study, along with previously discovered grouping effects based on color, shape, interpolated contours, and symmetry (Erlikhman et al., 2013; Feria, 2012; Howe & Holcombe, 2012b; Keane et al., 2011; Wang et al., 2016), may all represent concrete applications of Gestalt laws in multiple object tracking.

Gestalt laws mainly include the law of connectedness, law of similarity, law of closure, law of figure-ground, law of symmetry, law of common fate, etc. (Wertheimer, 1938). Based on the grouping effects discovered in multiple object tracking, it appears that grouping effects formed by different types of features correspond to specific Gestalt law principles. The target merging method used by Scholl et al. (2001), which connected two objects and caused them to be processed as a holistic object, is similar to the law of connectedness, where connected objects are grouped together. The grouping effects based on color, size, shape, contrast, stereo depth, or combined features discovered by Feria (2012), Howe and Holcombe (2012b), and Erlikhman et al. (2013) all obey the law of similarity. That is, objects with similar surface features are organized as a holistic unit.

The interpolated contour features discovered by Keane et al. (2011) and Erlikhman et al. (2013) conform to the law of closure. Several objects with closure characteristics are organized as a holistic unit. StClair et al. (2010) found that when texture motion direction is consistent with target motion direction, tracking performance is facilitated, whereas when it is opposite to target motion direction, tracking performance is impaired (StClair et al., 2010). This is similar to the law of figure–ground, where moving objects as a whole are perceived as one group and background texture as another group. Wang et al. (2016) found that when targets and targets are symmetric, tracking performance is better, whereas when targets and distractors are symmetric, tracking performance is worse. This is similar to the law of symmetry, where symmetric objects are organized together. Yantis (1992) found that relative speed differences between targets and distractors can improve tracking performance, even when this speed difference cannot be consciously perceived. Suganuma and Yokosawa (2006) found that when chase motion, parallel motion, or trail motion occurs between targets and distractors, tracking performance is reduced. These two studies demonstrate that common motion tendencies resulting from motion information such as speed, position, or direction all form grouping. This is consistent with the law of common fate.

The current study can also be viewed as a concrete manifestation of the law of common fate. When targets themselves have common fate internally, tracking is facilitated. Conversely, when targets and distractors have common fate, tracking is hindered. Therefore, Gestalt laws have important implications for studying grouping effects in multiple object tracking.

5.2 The Automatic Grouping Mechanism Based on Motion Information in Multiple Object Tracking

Regarding the automatic processing of grouping effects, Keane et al. (2011) first discovered that interpolated contour features affect multiple object tracking automatically, meaning that binding targets and distractors impairs multiple object tracking performance. Wang et al. (2016) found that regular features (symmetry) are also automatically processed in multiple object tracking; when targets and distractors are bound by symmetry rules in symmetric positions, tracking performance decreases. Erlikhman et al. (2013) comprehensively summarized various types of automatic surface features, temporarily dividing features affecting multiple object tracking into four categories. The first category is features irrelevant to tracking, such as orientation. However, this orientation is static, not motion direction. In fact, orientation is not completely irrelevant to tracking tasks either (Keane et al., 2011). The second category is voluntary processing features, such as stereo depth and contrast. These features can facilitate multiple object tracking performance when forming a group within targets, but do not impair performance when forming a group between targets and distractors. The third category is automatic non-grouping processing features, such as size. These features can automatically form grouping within targets but cannot automatically form grouping between targets and distractors. That is, when these features are shared between targets and distractors, they only disrupt the unity within targets but do not group targets and distractors sharing the same feature into one group. The fourth category is automatic grouping processing features, such as color, shape, feature combinations, and interpolation. The appearance of these features is automatically processed, whether appearing within targets or between targets and distractors. The common motion information in this study belongs to automatic grouping processing features, but its internal cognitive mechanisms have not been directly and fully elaborated in previous research. Therefore, there is currently no unified theory that can provide a definitive explanation for the automatic grouping mechanism based on motion information in multiple object tracking.

This study proposes that the automatic processing mechanism of motion information-based grouping effects needs to be elaborated by integrating Visual Index Theory, Grouping hypothesis, and visual working memory theory. First, according to Pylyshyn' s Visual Index Theory (Pylyshyn, 1989; Pylyshyn, 2001), in the pre-attentive stage, the visual index system pre-allocates a certain number of visual index tags to individuate visual objects to be tracked. That is, for the three motion conditions (target-target pairing, target-distractor pairing, and random no-pairing), each condition is allocated six visual index tags before motion begins. Visual index tags also have a certain stickiness; after being allocated to visual objects, they "stick" to the visual objects and can remain unchanged with environmental changes. Thereafter, when each object begins to move, according to Yantis' s (1992) Grouping hypothesis, observers consciously organize each tracking target into a higher-level virtual polygon during multiple object tracking. The vertices of this virtual polygon object correspond to each tracking target, meaning that the multiple object tracking process is essentially converted into a process of continuously processing a virtual polygon object with a constantly changing shape. Therefore, when object motion presents a paired chase state, the visual system automatically organizes front and rear moving objects into common-fate visual objects according to Gestalt perceptual organization principles. For example, when grouping forms within targets, front and rear targets have common fate due to common direction and same speed and are organized as common-fate visual objects. When grouping forms between targets and distractors, the front target and rear distractor are organized as common-fate visual objects due to common fate. When common-fate visual objects are formed, the motion information of front and rear objects is subsequently shared, and this motion information is stored in visual working memory. From the perspective of visual working memory, when targets and distractors share features, it leads to decreased multiple object tracking performance (Horowitz et al., 2007; Makovski & Jiang, 2009a, 2009b), but when targets and distractors do not share features, it facilitates multiple object tracking performance (Feria, 2012; Howe & Holcombe, 2012b). This is because sharing features between targets and distractors impairs feature independence, and this impairment is not caused by visual differences but originates from working memory confusion, damaging the unique identity information of visual objects and preventing lost objects from being recovered (Makovski & Jiang, 2009a, 2009b).

This is also similar to the situation where Scholl et al. (2001) used target merging methods to combine two visual targets. Although the merging methods differed, both resulted in visual object grouping and damaged the independence of visual index tags. Therefore, when grouping forms within targets, observers can rely on the position information of leading targets stored in visual working memory to reversely search for the position information of led targets. When grouping forms between targets and distractors, observers can no longer use the position information of leading targets stored in visual working memory for reverse search. In addition, the load of motion information stored in visual working memory may also differ between these two motion conditions. Ideally, in intra-target grouping, front and rear targets share motion information, meaning two targets can store the same motion information, whereas in target-distractor grouping, independent and different motion information needs to be stored for each target. That is, in terms of updating and maintaining spatiotemporal information, intra-target grouping may be superior to target-distractor grouping. Finally, the inhibition of distractors also differs. In the intra-target grouping condition, distractors do not share motion information with targets, so inhibiting distractors does not interfere with target memory information. Conversely, in the target-distractor grouping condition, distractors share motion information with targets, so inhibiting distractors interferes with target memory information.

In summary, it can be speculated that the multiple object tracking process based on motion information grouping first allocates visual index tags in the pre-attentive stage, then automatically groups according to Gestalt perceptual organization principles during motion to form visual objects with common fate. When grouping is formed, each target' s motion information is also stored in visual working memory space. When grouping forms within targets, targets and distractors do not share motion information, so observers can rely on common fate to retrieve lost targets, and the stored target motion information load is relatively reduced. When grouping forms between targets and distractors, targets and distractors share motion information, which prevents relying on common fate to retrieve lost targets, increases visual working memory load, and damages the visual uniqueness of each moving object. The above results only discuss the internal cognitive mechanism of automatic grouping based on motion information in multiple object tracking, while the involved cognitive neural mechanisms remain to be further explored.

5.3 The Influence of Surface Features on Motion Information-Based Grouping Effects in Multiple Object Tracking

Most previous studies have examined multiple object tracking performance under single features alone, such as Suganuma and Yokosawa (2006) exploring the influence of single spatiotemporal features on multiple object tracking. Makovski and Jiang (2009b) used color features as an operational variable to examine the role of unique identity working memory in multiple object tracking. Keane et al. (2011) discovered the automatic influence of interpolated contour features on multiple object tracking. Feria (2012) manipulated color feature similarity between targets and distractors to analyze how distractors affect multiple object tracking performance. Although Erlikhman et al. (2013) examined a large number of surface features, their analysis of each feature was independent, without much involvement of interactions between multiple features. Strictly speaking, although these studies only examined single features, from another perspective, these single surface features all caused changes in multiple object tracking performance. That is, surface features do affect the utilization of motion information in multiple object tracking. Fewer studies have involved interactions between features, such as Howe and Holcombe (2012b), who found by changing the combination of color and size features that when combined features do not share any feature with distractors, combined features can be used to facilitate multiple object tracking performance. Conversely, if they share a certain feature, combined features impair multiple object tracking performance. Wang et al. (2016) found additivity between regular features (symmetry) and surface features when examining regular feature-based grouping effects, but no additivity between two surface features. This suggests that interference may exist between features of the same type.

The current study examined the influence of color and shape features as surface features on motion information-based grouping effects and found that although the two surface features did not interfere with the formation of motion information-based grouping effects, their appearance significantly reduced motion information-based grouping effects. This implies that in multiple object tracking, motion information may belong to the same feature type as color and shape surface features. Furthermore, Suganuma and Yokosawa (2006) and Fencsik et al. (2007) found that motion information is stored in visual working memory, while Makovski and Jiang (2009b) found that surface feature information is also stored in visual working memory. Therefore, the cognitive mechanism by which surface features affect motion information grouping effects in multiple object tracking may originate from mutual occupation and interference between working memories.

6. Conclusion and Outlook

This study discovered an automatic grouping effect based on motion information, and this effect stably exists under different surface features (color and shape). This corresponds to the law of common fate in Gestalt principles. Additionally, although surface features do not interfere with the formation of motion information-based grouping effects, they significantly reduce the effect size of this grouping effect, suggesting that surface features and motion information may share visual working memory components.

Future research should further explore the neural mechanisms of motion information-based grouping effects to understand brain functional differences between spatiotemporal feature-based grouping effects and surface feature-based grouping effects. Simultaneously, more attention should be paid to the interaction relationships between different types of features in multiple object tracking. Finally, the role of Gestalt laws in multiple object tracking and their cognitive mechanisms can be further explored.

References

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433-436.

Erlikhman, G., Keane, B. P., Mettler, E., Horowitz, T. S., & Kellman, P. J. (2013). Automatic feature-based grouping during multiple object tracking. *Journal of Experimental Psychology: Human Perception & Performance*, 39(6), 1625–1637.

Fencsik, D. E., Horowitz, T. S., Kliege, S. B., & Wolf, J. M. (2004). Target reacquisition strategies in multiple object tracking. *Journal of Vision*, 4(8), 370–370.

Fencsik, D. E., Klieger, S. B., & Horowitz, T. S. (2007). The role of location and motion information in the tracking and recovery of moving objects. *Attention, Perception, & Psychophysics*, 69(4), 567–577.

Feria, C. S. (2012). The effects of distractors in multiple object tracking are modulated by the similarity of distractor and target features. *Perception*, 41(3), 287–304.

Feria, C. S. (2013). Speed has an effect on multiple object tracking independently of the number of close encounters between targets and distractors. *Attention, Perception, & Psychophysics*, 75(1), 53–67.

Franconeri, S. L., Pylyshyn, Z. W., & Scholl, B. J. (2012). A simple proximity heuristic allows tracking of multiple objects through occlusion. *Attention, Perception, & Psychophysics*, 74(4), 691–702.

Holcombe, A. O., Chen, W. Y., & Howe, P. D. (2014). Object tracking: Absence of long-range spatial interference supports resource theories. *Journal of Vision*, 14(6), 1-1.

Horowitz, T. S., Birnkrant, R. S., Fencsik, D. E., Tran, L., & Wolfe, J. M. (2006). How do we track invisible objects? *Psychonomic Bulletin & Review*, 13(3), 516–523.

Horowitz, T. S., & Cohen, M. A. (2010). Direction information in multiple object tracking is limited by a graded resource. *Attention, Perception & Psychophysics*, 72(7), 1765–1775.

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(2), 172-184.

Howe, P. D. L., & Holcombe, A. O. (2012a). Motion information is sometimes used as an aid to the visual tracking of objects. *Journal of Vision*, 12(13), 1-10.

Howe, P. D. L., & Holcombe, A. O. (2012b). The Effect of Visual Distinctiveness on Multiple Object Tracking Performance. *Frontiers in Psychology*, 3, 307-307.

Hu, L. M., Lyu, C., Zhang, X. M., & Wei, L. Q. (2018). The effect of identity switch in Multiple Identity Tracking. *Acta Psychologica Sinica*, 50(1), 9-27. [胡路明, 吕创, 张学民, 魏柳青. (2018). 身份交换对多身份追踪表现的影响. 心理学报, 50(1), 9-27.]

Huff, M., Meyerhoff, H. S., Papenmeier, F., & Jahn, G. (2010). Spatial updating of dynamic scenes: Tracking multiple invisible objects across viewpoint changes. *Attention, Perception, & Psychophysics*, 72(3), 628-636.

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(3), 685-698.

Keane, B. P., & Pylyshyn, Z. W. (2006). Is motion extrapolation employed in multiple object tracking? Tracking as a low-level, non-predictive function. *Cognitive Psychology*, 52(4), 346-368.

Lei, H. Y., Wei, L. Q., Lyu, C., Zhang, X. M., & Yan, X. Q. (2016). Expression-based Grouping in Multiple Identity Tracking. *Acta Psychologica Sinica*, 48, 141-152. [雷寰宇, 魏柳青, 吕创, 张学民, 闫晓倩. (2016). 多身份追踪中基于表情特征的分组效应. 心理学报, 48(2), 1-12.]

Luu, T., & Howe, P. D. (2015). Extrapolation occurs in multiple object tracking when eye movements are controlled. *Attention, Perception, & Psychophysics*, 77(6), 1919-1929.

Lyu, C., Hu, S., Wei, L., Zhang, X., & Talhelm, T. (2015). Brain Activation of Identity Switching in Multiple Identity Tracking Task. *PLOS ONE*, 10(12), e0145489.

Makovski, T., & Jiang, Y. V. (2009a). Feature binding in attentive tracking of distinct objects. *Visual Cognition*, 17(1-2), 180-194.

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(6), 1687-1697.

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(5), 631-671.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, 80(1-2), 127–158.

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

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, 32(1), 65–97.

Pylyshyn, Z. W. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual Cognition*, 11(7), 801–822.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking Multiple Items Through Occlusion: Clues to Visual Objecthood. *Cognitive Psychology*, 38(2), 259–290.

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*, 80(1-2), 159–177.

StClair, R., Huff, M., & Seiffert, A. E. (2010). Conflicting motion information impairs multiple object tracking. *Journal of Vision*, 10(4), 1–13.

Suganuma, M., & Yokosawa, K. (2006). Grouping and trajectory storage in multiple object tracking: Impairments due to common item motions. *Perception*, 35(4), 483–495.

Wang, C. D., Zhang, X. M., Li, Y. N., & Lyu, C. (2016). Additivity of Feature-based and Symmetry-based Grouping Effects in Multiple Object Tracking. *Frontiers in Psychology*, 7, 657–657.

Wei, L. Q., Zhang, X. M., Chuang, L., & Zhen, L. (2016). The Categorical Distinction Between Targets and Distractors Facilitates Tracking in Multiple Identity Tracking Task. *Frontiers in Psychology*, 7, 589–589.

Wertheimer, M. (1938). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 71–88). London, England: Kegan Paul, Trench, Trubner and Company.

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, 24(3), 295–340.

Zhang, X. M., Lu, X. M., & Wei, L. Q. (2011). The selective inhibition effects of changing the number of targets and nontargets on multiple object tracking. *Journal of Psychological Science*, 34(6), 1295–1301. [张学民, 鲁学明, 魏柳青. (2011). 目标与非目标数量变化对多目标追踪的选择性抑制效应. 心理科学, 34(6), 1295–1301.]

Zhang, X-M., Yao, L., & Lu, X-M. (2008). The role of location and trajectory information in the recovery of moving objects. *Acta Psychologica Sinica*, 40(1),

1-7. [张学民, 姚林, 鲁学明. (2008). 位置和轨迹信息在短暂消失目标复原过程中的作用. 心理学报, 40(1), 1-7.]

Note: Figure translations are in progress. See original paper for figures.