

Identity processing in multiple-face tracking

Dongning Ren

Institute of Psychology, Chinese Academy of Sciences,
Beijing, China, &
Graduate University, Chinese Academy of Sciences,
Beijing, China



Wenfeng Chen

State Key Laboratory of Brain and Cognitive Science,
Institute of Psychology, Chinese Academy of Sciences,
Beijing, China



Chang Hong Liu

Department of Psychology, University of Hull, Hull, UK



Xiaolan Fu

State Key Laboratory of Brain and Cognitive Science,
Institute of Psychology, Chinese Academy of Sciences,
Beijing, China



Research has shown that observers in a multiple-object tracking task are poor at recognizing the identity of successfully tracked objects (Z. W. Pylyshyn, 2004). Employing the same paradigm, we examined identity processing and its relationship with tracking performance for human faces. **Experiment 1** showed that although identity recognition was poorer after the target faces were learned in a dynamic display, identification performance was still much higher than the chance level. The experiment also found that on average about two face identities can be correctly traced to their locations. **Experiment 2** showed that tracking performance decreased significantly for unique upright faces relative to the unique inverted or identical upright faces, suggesting that upright faces activate some level of mandatory identity processing that interferes and competes with visual tracking for attentional resources. **Experiment 3** found that only target faces receive identity processing in the tracking task. **Experiment 4** showed that switching face identities during tracking impaired tracking performance. This may indicate that identity encoding is to some extent obligatory during multiple-face tracking. Furthermore, **Experiment 5** suggested that attentional resources can be consciously allocated either to maximize identity encoding or tracking, resulting in a tradeoff between the two. The results reveal a bias for face identity processing, which may differ significantly from multiple-object tracking.

Keywords: identity processing, multiple-face tracking, attention, mandatory process, resource sharing

Citation: Ren, D., Chen, W., Liu, C. H., & Fu, X. (2009). Identity processing in multiple-face tracking. *Journal of Vision*, 9(5):18, 1–15, <http://journalofvision.org/9/5/18/>, doi:10.1167/9.5.18.

Introduction

Research using the multiple-object tracking (MOT) paradigm has demonstrated that people are able to track four or more targets simultaneously (see Cavanagh & Alvarez, 2005, for a review). The finding is consistent with the observation that in a daily environment people are able to divide their attention among multiple regions of interest. The research has also generated less intuitive findings, particularly concerning the issue of dynamic binding of identity and spatiotemporal information. Objects are commonly distinguished by features. Yet relative to object locations, object features such as color and shape are often lost easily during tracking (Bahrami, 2003; Saiki, 2003a, 2003b). Intuitively, individuating objects by colors should facilitate tracking. However, this does not always have an advantage over tracking physically identical objects (Klieger, Horowitz, & Wolfe,

2004). Although latest evidence does show that unique object shapes facilitate tracking, identity recognition remains worse than tracking performance (Horowitz et al., 2007). Consistent with this poor binding of object features and location, research has found that people are poor at identifying correctly tracked objects (Pylyshyn, 2004). This seems rather puzzling because successful tracking requires correct target tagging. To explain this paradoxical phenomenon, Pylyshyn suggests that multiple-object tracking may be implemented by low-level vision, where the information about individual identity is encapsulated and inaccessible from higher level cognition.

While identity–location binding often has little consequence in multiple-object tracking, it can be vital in social interactions. It is the responsibility of a caretaker to scrutinize the activities of children in the playground, noting the specific events linked with each individual. An eyewitness of a crime scene has to correctly associate the acts and locations with each involved suspect although

this is by no means an easy task. A key motivation behind MOT studies is to understand how identity and spatio-temporal events are bound together in this kind of situations. To address this issue, we employ the MOT paradigm to examine the facial identity and location binding in this study. Although every part of the human body carries useful clues to an identity, the human face is arguably the most salient source of information for person identification. The principal objective of this study is to investigate whether processing of facial identity is to some extent obligatory in multiple-face tracking and how attentional resources are used for identity processing and location tracking. It is difficult to infer the relationship between tracking and processing of facial identity from the existing MOT literature because most studies to date have employed nonface stimuli such as simple geometric shapes and line drawings of objects. Identity processing for faces and objects may not be the same. Research has shown that face identification relies on both featural and configural information, and there are notable differences between entry-level object recognition and face recognition (Bruce & Humphreys, 1994).

To our knowledge, only a recent study by Oksama and Hyönä (2008) has employed face stimuli in MOT. The main purpose of their study was quite different from ours. They examined how familiar/famous faces affect tracking performance and found that familiar faces were easier to track than pseudo-faces. Our study, on the other hand, used unfamiliar faces as stimuli. There is evidence that identity processing for familiar and unfamiliar faces demands different level of attentional resources (Jackson & Raymond, 2006). The main difference, however, is that our study focuses on whether identity processing of unfamiliar faces is to some extent spontaneous or mandatory without deliberate intentions and whether voluntary control modulates the outcome of dynamic binding of location and identity. In addition, instead of a small number of line drawings, we employed a large number of photographic images to improve the chance of generalization. Line drawings may be limited for understanding face processing because they lack reflectance cues and surface information that are important for face recognition (Bruce, Hanna, Dench, Healey, & Burton, 1992; Russell, Biederman, Nederhouser, & Sinha, 2007; Vuong, Peissig, Harrison, & Tarr, 2005). Line drawings are also known to impair configural processing in face recognition (Leder, 1996).

Although the role of attention in processing facial identity has been studied extensively in recent years, no study to date has employed the MOT paradigm for this purpose. However, the paradigm has obvious advantages for the study of attention in face processing. In reality, the location of a face is rarely fixed. Moreover, it is often necessary to achieve dynamic binding of multiple faces and locations. The role of attention in multiple-face tracking is clearly important for understanding social interactions. Because the current theories would make

different predictions about the role of attention in this type of tasks, we first briefly outline some main theoretical positions.

The role of attention in processing of facial identity

Research has shown that faces attract more attention than objects or other environmental stimuli. Certain facial information such as expressions of anger and fear are processed automatically in an obligatory fashion (see a review by Palermo & Rhodes, 2007). The role of attention in processing of facial identity is more controversial. Jackson and Raymond (2006) summarized three theories in detail. Here we only give a brief sketch of each.

According to the first theory, attention is not needed, because irrelevant face distractors can interfere with a name-categorization task (Young, Ellis, Flude, McWeeny, & Hay, 1986). Identity information appears to be processed even when attention was directed away from face distractors by a target with high perceptual load (Lavie, Ro, & Russell, 2003). Configural processing in face recognition is considered automatic because the identity of a face can be coded even when participants are told to ignore it (Boutet, Gentes-Hawn, & Chaudhuri, 2002).

The second theory proposes that processing of facial identity has access to a separate attentional mechanism. The evidence for this comes from a series of dual-task experiments (Palermo & Rhodes, 2002), which showed no task interference when one task involved matching facial features whereas the other involved matching facial configurations. Interference occurred when both tasks involved configural processing. Similar findings were reported by Awh et al. (2004). This leads to the idea that separate attentional resources are available for featural and configural processing in face recognition.

The third theory argues that face identification requires the same kind of attention as object recognition (Downing, Liu, & Kanwisher, 2001; Wojciulik, Kanwisher, & Driver, 1998). Face processing in visual search tasks seems to require effortful attention because faces do not produce pop-out effects (Brown, Huey, & Findlay, 1997; Kuehn & Jolicoeur, 1994; Nothdurft, 1993). Experiments using the attentional blink procedure also found that face identification can be impaired if attentional resources are temporarily occupied by another concurrent identification task for a nonface pattern (Jackson & Raymond, 2006). The results are consistent with the attentional blink effect found for nonface stimuli. Therefore, the same accounts of the effect appear to apply in both cases.

Recently, research using a variant of target–distractor interference paradigm developed by Young et al. (1986) has provided evidence that no more than a single facial identity can be processed at a time (Bindemann, Burton, & Jenkins, 2005). Because a face–face congruency effect

was not found, it led to the interpretation that the capacity limit for face processing is one face.

The literature has thus shown a wide spectrum of theories on the role of attention in processing of facial identity. They make different predictions for face identification in a multiple-face tracking paradigm. First, if no attention is required for face identification, multiple-face tracking should not affect recognition of either target or distractor faces because face processing will not have to compete with tracking for attentional resources. Second, if face identity processing engages separate attentional resources, tracking performance should not be affected, but only a small number of faces can be processed serially, due to the limits of resource for configural processing. Third, if the same attentional sources are used for both tracking and face identification, there should be a tradeoff between identity and tracking processing because of the resource sharing. Finally, face identification performance should be poor if only one facial identity can be processed at a time.

To examine these predictions, we measured the effects of attentional load due to explicit or implicit face encoding. Our main purpose was to examine whether identity information could be retained after multiple-face tracking and whether facial processing of identity is to some extent mandatory.

Identity processing in multiple-object tracking

The literature on MOT has also generated different views on identity processing. A detailed review of this can be found in Horowitz et al. (2007). Here we only briefly summarize two most relevant ones. A key issue in this literature is whether object identity is content addressable. In his Fingers of INSTantiation (FINST) model, Pylyshyn (1989, 2001) suggests that MOT is implemented by early vision that picks out a small number of objects while ignoring their visual properties. According to this theory, the object identity differentiated by visual properties is not encoded or accessible from higher level cognitive processes even when the objects that have those properties are attended. This idea has been supported by the evidence where observers were unable to report the features or identity of the objects used in MOT tasks (Pylyshyn, 2004; Scholl, Pylyshyn, & Franconeri, 1999). An important feature of this model is that the tracking mechanism is data-driven and preattentive.

In contrast to FINST, the object file theory assumes that featural properties along with other semantic information of the object are encoded and updated through time and space (Kahneman, Treisman, & Gibbs, 1992). A prominent feature of this theory is that object files are content addressable. Evidence for this theory can be found in Horowitz et al. (2007) and Oksama and Hyönä (2004, 2008) who demonstrated that their observers were able to identify some of the tracked objects.

These theories make different predictions about the outcome of multiple-face tracking in this study. If tracking is unaffected by visual properties of the tracked items as the FINST model suggests, the presence of multiple unique or identical faces should not affect tracking performance. If identity information is available as the object file model suggests, some identity information should be preserved and retrievable after multiple-face tracking.

Experiment 1

The purpose of this experiment was to measure the effect of multiple-face tracking on identity processing. To achieve this, [Experiment 1A](#) compared recognition performance either after tracking or after viewing static face images, whereas [Experiment 1B](#) investigated whether a face could be correctly linked to its location after multiple-face tracking.

Experiment 1A

If processing of facial identity requires no attention, then multiple-face tracking should have little effect on the subsequent face identification performance. If identity information is not encoded, then face identification should be nearly impossible after tracking. Although participants in the dynamic condition of this experiment had to track the target faces, the task was different from typical MOT tasks because the participants were only told to remember the target faces without indicating the target locations. At the end of each trial, participants performed a 3-alternative-forced-choice task (3AFC) in which they had to identify the target face embedded in two distractors.

Methods

Participants

Ten undergraduate students from Chinese Agriculture University participated in this experiment for a small payment. All had normal or corrected-to-normal vision. Participants in all experiments were naive about the purpose of the study.

Stimuli

The stimuli were selected from the CAS-PEAL Large-Scale Chinese Face Database (Gao et al., 2008). They were grayscale images of frontal view faces with neutral expression. A total of 792 faces was used in this experiment, including 468 males and 324 females. The images were cropped to remove shoulders and extraneous

background. The size of the faces was normalized according to the face width. The resulting image size was 60×75 pixels, which measured 2.4×3.0 cm ($2.3 \times 2.8^\circ$) on screen. All images were scaled to the same mean luminance and root-mean-square contrast.

The stimuli were presented on a 17" Lenovo CRT monitor with the refresh rate of 100 Hz. A central square area of 480×480 pixels, subtended $18.2 \times 18.2^\circ$ of visual angle at a viewing distance of 60 cm, was designated for stimulus presentation. The background color of the display was a homogenous gray. We used E-Prime (Version 1.1) to generate the dynamic tracking and still displays and to control the flow of the experiment.

Design and procedure

We employed a 2×3 within-participant design. The two variables were the learning condition (tracking vs. still) and the number of targets (1, 3, or 4 faces). We did not include 2 targets as a condition because the results for 1 and 2 targets were indistinguishable in a pilot test. The single-target condition was included as a baseline for multiple-face tracking, which by definition involves at least two or more targets.

The two learning conditions were run in two blocks of trials. The targets and distractors either moved about the screen or remained still in these blocks. Each block began with two practice trials followed by 42 experimental trials (14 trials \times 3 levels of target quantity). Half of the participants completed the face tracking block first whereas the rest completed the still-image block first. The order for the number of target faces within a block was random.

The faces presented in each trial were of the same sex. They were randomly chosen for each participant from the

pool of 792 faces. The faces presented in every trial were different. Namely, once a face was assigned to one trial, it would not be used again in the subsequent trials. The procedure in each trial is illustrated in Figure 1A. It began with 8 black rectangles on the screen. The location of the rectangles was randomly assigned, with the constraint that none would occlude the others. A subset of the rectangles would then start to blink twice for 2 s, signaling the target location. Following this, the rectangles changed abruptly into 8 moving faces, which moved in random directions. The faces bounced off each other when the center-to-center distance was less than twice of their size. They avoided the edge of the display area. The velocity of the face images varied between 4.2 and $7.4^\circ/s$ with a mean of $5.8^\circ/s$. Each face changed its speed and direction randomly at each frame. The change occurred every 500 ms on average. The speed was randomly selected between 4 and 8 pixels per frame.

The duration of the dynamic display was proportional to the number of targets. We gave 1 s for each target face. The duration for 1, 3, and 4 targets was thus 1, 3, and 4 s, respectively. The screen was cleared at the end followed by a 1-s blank screen. Then 3 faces were presented in a row in the center of the screen. These consisted of a target, a distractor, and a new face that was not shown during learning. The new face was used to estimate whether participants also processed the identity of the old distractor face during target tracking. In case there were several targets, only one of them was chosen randomly. The task of the participant was to click on the target face with a mouse. No feedback was given. The next trial was initiated by the participant pressing the space bar.

In the still-image condition, the initial trial sequence was identical to the tracking condition except that the faces remained stationary after replacing the rectangles.

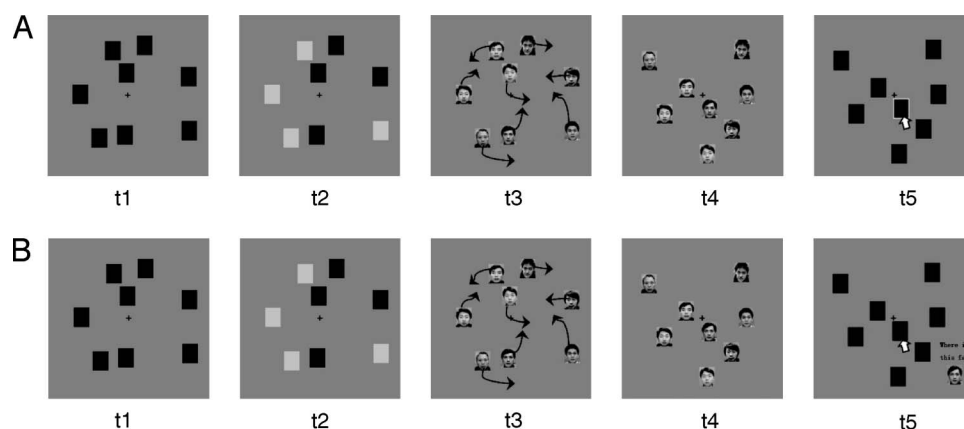


Figure 1. (A) Illustration of the trial procedure used in Experiment 1A. Targets are marked at t2. Tracking commences at t3 and stops at t4, when the faces are immediately masked by a blank screen. The task is to identify the target face at t5. (B) Illustration of the trial procedure used in Experiment 1B, the procedure from t1 to t3 is identical to that used in Experiment 1A. At t4, immediately after tracking stops, the faces are occluded by rectangles. For the specific task, the observer has to indicate the location of a randomly selected target face by a mouse click on a rectangle at t5, as illustrated in this example. For the standard task, no probe face was shown. The observer only has to click on all the target locations without linking each target face to a unique location.

Results

The mean percent correct responses are shown in Figure 2. A repeated-measures analysis of variance (ANOVA) showed that the recognition performance for the still condition was better than the dynamic condition, $F(1, 9) = 6.30, p < 0.05$. There was also a significant main effect for number of targets, $F(2, 18) = 22.94, p < 0.001$. Performance for learning one face was better than for learning 3 or 4 faces ($ps < 0.01$). The difference between learning 3 and 4 faces was not significant ($p = 0.31$). The interaction between the two variables was also not significant, $F(2, 18) = 1.23, p = 0.32$.

To find out whether the distractors were more likely to be mistaken as targets than the new faces, we performed a $2 \times 2 \times 3$ ANOVA on the false alarm results, adding the two types of distractor as another variable. The results (see Table 1) showed that learning moving targets produced higher false alarms than learning still targets, $F(1, 9) = 6.26, p < 0.03$. However, there was no difference between the results of distractors and new faces, $F(1, 9) = 1.02, p = 0.34$, or interactions between the two factors, $F(1, 9) = 0, p = 1.00$. Other main effects or interactions were also not significant.

Discussion

Although the recognition performance was worse when the target and distractor faces moved around the screen than when they were still, even the lowest accuracy rate for 4 moving target faces was 74%. This was much higher than the 33% chance level. It shows that a significant amount of identity information can be processed and retained in multiple-face tracking. The result is different from Pylyshyn (2004)'s observation that object identity cannot be accessed explicitly. However, there was a clear cost of tracking compared to the still condition. This supports the idea that the attentional resources demanded by tracking can damage identity processing. The results are consistent with Oksama and Hyönä (2004, 2008) and

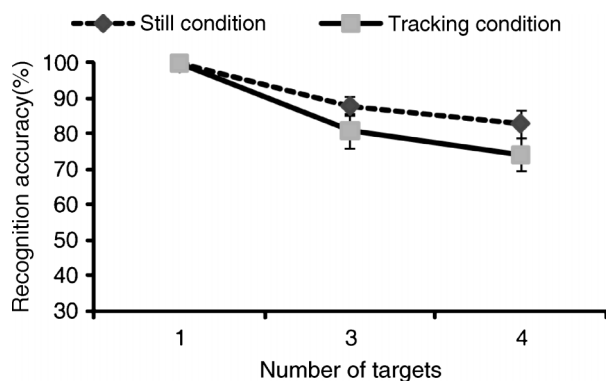


Figure 2. Mean accuracy results in Experiment 1A. The error bars represent one standard error.

	Stimulus category	Number of targets		
		1	3	4
Still	Distractor	0 (0)	6.4 (5.3)	10.0 (9.0)
	New	0 (0)	5.7 (4.5)	7.1 (5.8)
Tracking	Distractor	0 (0)	10.0 (9.6)	14.3 (11.2)
	New	0 (0)	9.3 (10.1)	11.4 (9.0)

Table 1. False alarm rate (%) of distractor and new faces. Values in parentheses are standard deviations.

Horowitz et al. (2007) who observed similar dynamic binding of identity and location in multiple-face/object tracking.

Experiment 1B

Although Experiment 1A demonstrated that observers could report the identity of tracked objects, it does not show that there is a link between identity and location. To make a stronger case for content addressability, we employed a task described in Horowitz et al. (2007) that required linking identity to location. If the representations of the tracked faces are content addressable, then the observer should be able to specify the exact location for a randomly probed target after multiple-face tracking.

Methods

Participants

A different group of eight undergraduate students from Chinese Agriculture University participated in this experiment for a small payment. All had normal or corrected-to-normal vision.

Stimuli

The stimuli and apparatus were identical to Experiment 1A.

Design and procedure

Two task conditions were employed in this experiment. In one condition, the observer was only required to indicate the locations for all targets without specifying which target face was presented in which condition. This is the typical MOT procedure, which is referred to as the standard task. In the other condition, one of the faces was randomly chosen from the targets and probed at the end of each trial. The observer had to specify the exact location of the probe face. This is a type of partial report procedure, which is referred to as the specific task. Both tracking tasks consisted of either 3 or 4 target faces.

The task conditions were run in two blocks of trials. The order of the blocks was counterbalanced across

participants. Each block began with four practice trials followed by 30 experimental trials (15 trials \times 2 levels of target quantity).

As illustrated in [Figure 1B](#), the procedure in each trial was similar to the tracking task in [Experiment 1A](#), except that immediately after motion stopped, all the faces were occluded by rectangles and the participant was either required to indicate the locations of all targets by clicking on the rectangles (standard task) or the specific location of a randomly probed target face (specific task). The selected rectangle was highlighted by a yellow border, which could be turned on and off by clicking. This allowed the participant to change the answer before starting the next trial via a key press. No feedback was given.

Results

Following the equation detailed in Horowitz et al. (2007), a common metric of capacity k was computed from the number of faces tracked corrected for guessing:

$$k = \frac{a - pt - 1 - \sqrt{(1 - a - pt)^2 - 4(apt - t)}}{2}, \quad (1)$$

where a is the number of possible response options, p is the performance in terms of the number of targets correctly identified, and t is the number of targets.

This allowed for comparisons of the results between the two task conditions. The mean capacity results are shown in [Figure 3](#). The results indicate that the observers were able to identify the exact locations for up to about two items in the specific condition, $ts(7) > 8.01$, $ps < 0.01$. There was no difference between capacities for 3 or 4 targets, $t(7) = 0.19$, $p = 0.86$. Estimated capacity was lower in the specific condition than in the standard condition for both 3 and 4 targets, $ts(7) > 4.90$, $ps < 0.01$.

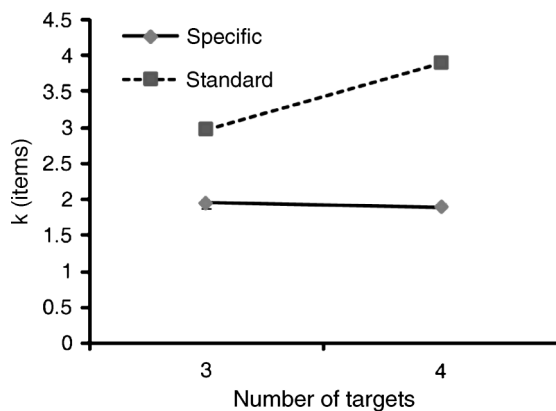


Figure 3. Mean capacity results in [Experiment 1B](#). The error bars (all nearly zero) represent one standard error.

Discussion

[Experiment 1B](#) provided further evidence that representations of targets are content addressable although the finding suggests that the capacity of identity processing may be limited to about two items. The overall results are consistent with the findings in Horowitz et al. (2007).

Experiment 2

Although tracking was an implicit task for the participants in [Experiment 1A](#) because they were not explicitly required to do so, it was nevertheless necessary for the targets to be encoded correctly. Like prior studies (Bahrami, 2003; Oksama & Hyönä, 2008; Scholl et al., 1999), the explicit task in [Experiment 1A](#) was to learn faces. It was used to examine the extent to which attention for processing of facial identity can be consciously modulated. However, this explicit demand creates the possibility that identity processing in multiple-face tracking requires conscious efforts to allocate attentional resources. Our next question is therefore whether identity processing could occur without explicit intention. This was explored in [Experiment 2](#). Unlike [Experiment 1](#), participants in this experiment were not required to learn the identity of the faces. Instead, they were only required to perform the tracking task. If identity information were not encoded in multiple-face tracking unless participants were explicitly instructed to do so, then tracking different or identical faces should make no difference to tracking performance. Alternatively, if identity information were processed even when it was task irrelevant, then tracking different faces should produce poorer performance than tracking identical ones because extra attentional resources would be taken away from tracking for face encoding.

To test these alternative hypotheses, we compared the tracking performance for different faces and identical faces. We also included a condition where the different faces were presented upside down. Although the inverted faces had exactly the same physical features as the upright faces, inversion is known to impair face encoding and recognition (see reviews by Farah, Wilson, Drain, & Tanaka, 1998; Searcy & Bartlett, 1996). We therefore predicted that when different faces are inverted, they should produce different tracking performance relative to upright different faces.

Methods

Participants

A different group of nine undergraduate students from the Chinese Agriculture University participated in this experiment for a small payment. All had normal or corrected-to-normal vision.

Stimuli

The stimuli and apparatus were identical to [Experiment 1A](#), except that the image size was reduced from 2.4×3.0 cm to 2.0×2.5 cm to create room for more faces in the same display area. The number of faces was increased from 8 to 10 because a pilot study showed a ceiling effect for 8 faces (above 98%). We used a total of 671 faces in this experiment including 367 males and 304 females.

Design and procedure

We employed a 2×3 within-participant design. The two variables were stimulus type (different upright faces, identical upright faces, and different inverted faces) and the number of target faces (3 or 5).

The tracking task was identical to [Experiment 1A](#) except for the following. The identical face condition was created by randomly selecting a face from the pool of face stimuli. The same face was then used for all the targets and distractors in that trial. There was no identification test at the end of each trial. There was only one block of 90 experimental trials after 5 practice trials. There were 15 trials for each of the six conditions. The order of 90 trials for these six conditions was random. There were 10 faces (instead of 8) in each trial. Among these, 3 or 5 faces were randomly assigned as targets. The face stimuli moved about the display for 7 s in both target conditions. The mean velocity of the item movement was $5.7^\circ/\text{s}$. The task of the observers was to indicate which faces were the targets.

Results

[Figure 4](#) shows the tracking accuracy. ANOVA revealed significant main effects of stimulus type, $F(2, 16) = 7.26$, $p < 0.01$, and number of targets, $F(1, 8) = 7.64$, $p < 0.05$, where 3 targets were tracked better than 5 targets. There was also a significant interaction, $F(2, 16) = 4.18$, $p < 0.05$. Simple main effects analyses revealed that the

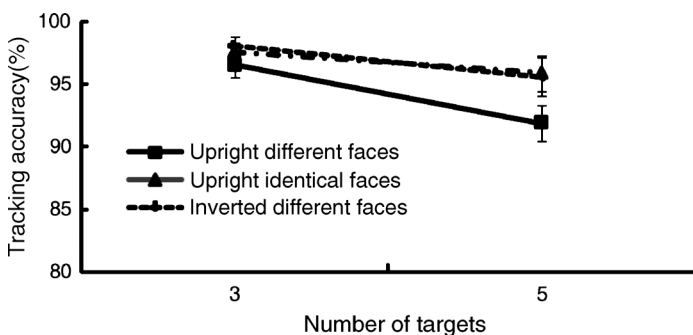


Figure 4. Tracking accuracy in [Experiment 2](#). The error bars represent one standard error.

difference among stimulus types was not significant for the 3 target condition, $F(2, 16) = 0.84$, $p = 0.45$, but was for the 5 target condition, $F(2, 16) = 15.30$, $p < 0.001$, where performance for upright different faces was worse than both the upright identical faces and inverted different faces, $ps < 0.01$. The results for upright identical faces and inverted different faces were the same, $p = 1.00$.

Discussion

When participants tracked 3 targets, results from the 3 types of stimuli were similar. This may have been due to a ceiling effect. As targets increased to 5, performance for different upright faces fell more sharply than identical upright and different inverted faces. This suggests that upright faces engage attentional resources for identity processing to some extent that in turn impairs tracking performance. The identical upright face condition may have involved less identity processing, because low-level image processing could quickly confirm that all images used in this condition were identical. In contrast to [Horowitz et al. \(2007\)](#) and [Klieger et al. \(2004\)](#), where different objects either produced equivalent or better tracking performance than identical objects, our results show that the benefit of using multiple unique items in tracking can be reversed when face stimuli are employed. Results from this experiment suggest that the identity of upright faces may be processed even when identity processing is task irrelevant. However, the results cannot tell us whether identities of target and distractor faces were equally processed. The next experiment was designed to address this question.

Experiment 3

Results in [Experiment 1A](#) suggest that distractor faces may not be processed because the distractors that were shown with the target faces during the learning phase did not create more false alarms than the new faces that were not shown during learning. However, this could be because participants in [Experiment 1A](#) were explicitly instructed to learn target faces. In [Experiment 2](#), although there were signs of identity processing for different upright faces, it was not possible to tell whether target and distractor identities were equally processed.

In the MOT literature, there is evidence that target identities are detected more accurately than distractors ([Bahrami, 2003](#)). However, because the results were mainly based on nonface stimuli, it is difficult to ascertain whether they can be generalized to faces. If processing of facial identity is to some extent mandatory for both targets and distractors, recognition performance for these should be similar. However, if attention is primarily directed to

the targets during tracking, the identity of distractors may not be encoded equally well.

To test these hypotheses, each trial in [Experiment 3](#) was followed by a tracking test and an identification test. The tracking test was the same as in [Experiment 2](#). The identification test required participants to judge whether 8 sequentially presented faces were seen in the tracking phase. Some of these faces were selected from targets, and others from distractors and new faces. Participants were told that the two tasks were unrelated. This was used to eliminate the potential demanding characteristics of the experiment for exclusive target identity processing.

Methods

Participants

Ten different undergraduate students from the Chinese Agriculture University were paid for their participation. All had normal or corrected-to-normal vision.

Stimuli

Apparatus and stimuli were identical to [Experiment 1A](#). A total of 740 faces was used, with the same number of males and females.

Design and procedure

This was a within-participant design. The tracking phase was the same as [Experiment 1A](#) except for the following. The number of targets was always 4. The face images moved about the screen for 10 s in all trials. The velocity for each face varied between 4.2 and 7.4°/s with a mean velocity of 6.1°/s. Participants in this experiment were required to identify the 4 target locations. Following this, participants were shown 8 faces, which consisted of 4 target faces, 2 distractor faces that were used in the tracking phase, and 2 new faces that were not used in the tracking phase. The order of these was random. These test faces were presented one at a time in the center of the screen for 300 ms followed by a blank screen. Participants

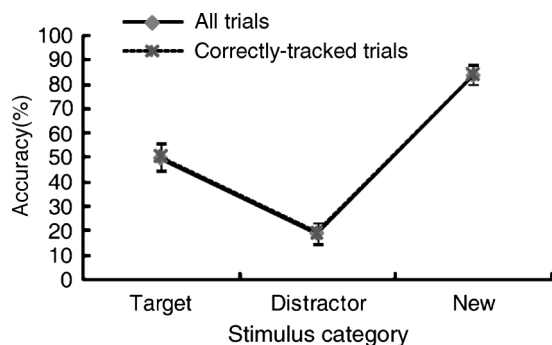


Figure 5. Accuracy results in [Experiment 3](#). The error bars represent one standard error.

	Response	Stimulus category		
		Target	Distractor	New
All trials	Yes	684 (88)	657 (107)	649 (71)
	No	670 (70)	650 (48)	645 (53)
Correctly tracked trials	Yes	664 (94)	617 (66)	654 (100)
	No	664 (66)	649 (51)	640 (52)

Table 2. Mean reaction time results (ms) in [Experiment 3](#). Values in parentheses are standard deviations.

were asked to judge whether the face was shown in the tracking phase as quickly as possible. Because some of the target and distractor faces used in the tracking phase were shown again in the identification task, we used the reaction time to measure whether this could produce a repetition priming effect. There were 4 practice trials followed by 70 experimental trials.

Results

The overall tracking accuracy was 95%. In 82% of the trials, the tracking performance was 100% correct where all 4 targets were tracked correctly.

The results of recognition performance are shown in [Figure 5](#). We conducted two separate ANOVAs, one for all trials and the other for trials where the tracking performance contained no errors. A significant main effect was found in both analyses, $F_s(2, 18) = 39.30$ and 38.51 , respectively, $p_s < 0.001$. Post-hoc comparison of means in both analyses showed a better performance for the new faces than for the targets or distractors ($p_s < 0.01$). More importantly, the target faces produced higher accuracy than the distractors ($p < 0.01$).

The reaction time results are shown in [Table 2](#). A 2×3 ANOVA revealed no significant main effects and interaction (all $F_s < 3$, $p_s > 0.15$).

Discussion

The results show that the participants were able to discriminate the new faces from the target and the distractor faces. It indicates that the identity of the faces was processed in the tracking task. More importantly, target faces were over twice more likely to be identified than distractor faces. This result is consistent with [Bahrami \(2003\)](#), who found that color and shape changes in targets are easier to detect in MOT. Both results are different from [Scholl et al.'s \(1999\)](#) report that identity information is equally unavailable for tracked and untracked items.

The results in this experiment suggest that the impaired tracking performance found in the different upright face condition of [Experiment 2](#) was likely due to processing of

the target identity. However, although this may have occurred involuntarily, it is also possible that the identity processing was a result of voluntary strategy. Participants might have attempted to encode the unique face identities believing it would help them with the tracking task, although this might be an ineffective strategy. We examined this alternative explanation in [Experiment 4](#).

Experiment 4

To determine whether the identity processing in the previous experiments could be explained by a voluntary strategy for a better tracking performance, we compared two conditions in this experiment, where the identity of each face either remained the same or switched from time to time to a different person during the course of tracking. The identity switching manipulation was expected to prevent the voluntary strategy because there would be no reliable relationship between facial identities and the tracked items. If the impaired tracking performance in the previous experiments was due to involuntary face processing taking resources away from tracking, this manipulation should either create the same result as in the standard condition, or even amplify the costs to tracking, because face identity processing would be strained by the manipulation and result in higher resource demands.

Methods

Participants

A different group of twelve paid participants was recruited from the Chinese Agriculture University. All had normal or corrected-to-normal vision.

Stimuli

These were identical to [Experiment 2](#).

Design and procedure

The two tracking conditions were tested in a block design, where the order of the blocks was counterbalanced across participants. The number of target faces in this experiment was always five. In one tracking condition, the identity of each face switched every 500 ms from one to another with the constraint that each face must change identity. The other tracking condition was the same as [Experiment 2](#).

Results

[Figure 6](#) shows the results of tracking. The results show that changing the face identities during tracking signifi-

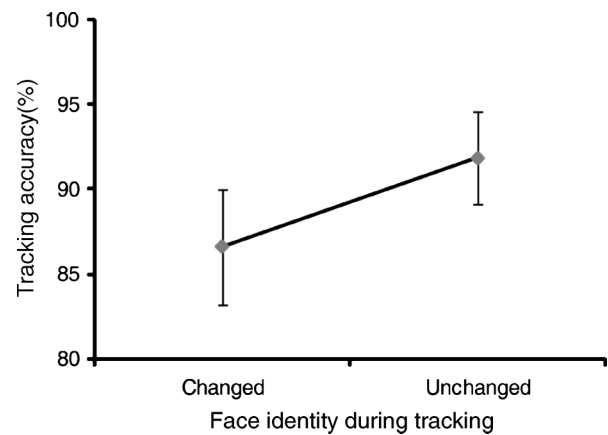


Figure 6. Tracking accuracy in [Experiment 4](#). The error bars represent one standard error.

cantly impaired the tracking performance relative to the condition where no change of identities occurred during tracking, $F(1, 11) = 9.78$, $p < 0.01$.

Discussion

If the cost of identity processing found in our previous experiments was due to an ineffective voluntary strategy, the tracking performance should be improved when potential identity cues for this strategy are removed. Contrary to this prediction, tracking performance in this experiment worsened when the reliable link between facial identities and tracked items was severed by identity switching. The result is consistent with the hypothesis that the impaired tracking performance was due to some level of involuntary face identity processing.

This experiment shows that identity encoding is to some extent mandatory. However, this does not rule out the possibility that conscious efforts to learn the target faces could further impair tracking performance. We examined the effect of effortful identity learning in [Experiment 5](#).

Experiment 5

We can infer from the results of [Experiment 2](#) that using different upright faces in multiple-face tracking can impair tracking performance. [Experiments 3](#) and [4](#) suggest that the interference may reflect involuntary processing of target identity even when it is task irrelevant. To what extent the attentional resources can be voluntarily allocated to tracking and face identification required in [Experiments 1](#) and [3](#)? Would tracking performance be affected further if target face identity is consciously

pursued and encoded? To answer these questions, we compared tracking performance in two conditions. Half of our participants in this experiment performed the tracking task only, whereas the other half also performed a face identification task after the tracking task. If attentional resources could be assigned voluntarily, we would expect an impaired tracking performance in the second group where attempts would be made to maximize performance for both tracking and identification tasks.

Methods

Participants

A different group of 20 paid participants was recruited from the Chinese Agriculture University. All had normal or corrected-to-normal vision. They were randomly assigned to the two conditions, with 10 participants in each.

Stimuli

These were identical to [Experiment 3](#).

Design and procedure

We employed a between-participant design for this experiment. The two conditions were tracking only and tracking and learning.

Participants in both groups performed the same tracking task as in [Experiment 3](#). After this, participants in the tracking-and-learning condition also performed an identification task. In tracking-and-learning condition, participants were explicitly instructed to learn the faces during tracking. Unlike [Experiment 3](#) where participants were expected to identify both targets and distractors presented in the tracking phase, the identification task in this experiment only required learning the target faces. We focused on the processing of target identity here because [Experiment 3](#) showed that distractor identity is unlikely to interfere with the tracking task. In addition, we did not require timed response in this experiment because [Experiment 3](#) showed that the reaction time data were less sensitive to experimental manipulations than the accuracy data. Participants in the tracking-only condition simply viewed these faces without having to perform the identification task. Other aspects of the procedure for the identification task were identical to [Experiment 3](#).

Results

The tracking performance for the tracking-and-learning condition ($M = 90\%$, $SD = 4$) was lower than tracking-

only condition ($M = 97\%$, $SD = 3$), $t(18) = 2.84$, $p < 0.01$.

The accuracy results of the face identification task in the tracking-and-learning condition are shown in [Figure 7](#). It should be noted that because the task in this condition was to learn target faces, the correct response to a distractor face was “no”. This was different from [Experiment 3](#) where the correct answer was “yes” because the task required learning the distractor faces as well. We performed two one-way ANOVAs, one for results from all trials, and the other for results from the 68% of trials where no tracking mistakes were made. A significant main effect was found in both analyses, $F_s(2, 18) = 28.12$ and 25.57 , respectively, $p_s < 0.001$. Post-hoc tests in both analyses confirmed identical performance for the distractor and new faces ($p_s = 0.49$ and 0.97 , respectively), although accuracies for both distractor and new faces were higher than the target faces ($p_s < 0.01$).

Discussion

Results in this experiment demonstrate that tracking performance suffers when voluntary effort is made to learn target facial identity. This was not only shown by the direct comparison of the tracking performance for the tracking-only and tracking-and-learning conditions but was also indirectly suggested by a drop of the error-free tracking from 82% of the trials in [Experiment 3](#) to 68% in the tracking-and-learning condition of this experiment. Due to the emphasis on the speed of response in [Experiment 3](#), more resources might have been available for a better tracking performance. This experiment shows that in addition to the effect of mandatory identity processing, tracking performance can deteriorate further if there is an explicit intention to learn the identity of the faces.

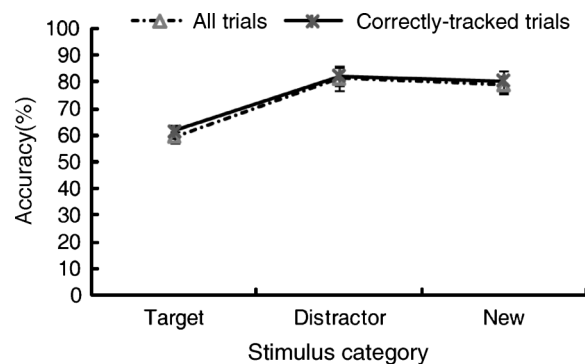


Figure 7. Recognition accuracy in the tracking-and-learning condition of [Experiment 5](#). The error bars represent one standard error.

The lack of difference between old and new distractors in the recognition performance suggests that distractor facial identity in the tracking phase is not processed. This result replicates the finding in [Experiment 1](#).

General discussion

We conducted five experiments to examine location–identity binding and the role of attention in multiple-face tracking. Results in [Experiment 1](#) showed that faces are more difficult to learn if they constantly change locations during learning. However, despite this cost of multiple target tracking, face identification performance was much better than the chance level. It indicates that processing of facial identity cannot be abolished because of tracking. The experiment also showed that multiple face tracking is to some extent content addressable because at least some of the tracked faces can be correctly linked to their locations.

Unlike [Experiment 1](#) where the sole explicit goal was to learn the target identity, [Experiment 2](#) examined tracking performance when participants were not required to process the identity of the tracked items. The faces in each trial were either upright images of the same individual, upright images of different individuals, or inverted images of different individuals. It was found that tracking for the upright different faces was poorer than the upright identical or different inverted faces. This impairment for upright faces may be caused by voluntary or mandatory face identity processing.

[Experiment 3](#) explored whether this identity processing happened to both target and distractor faces. The results showed that target faces were more likely to be encoded during tracking. To determine whether identity processing for upright target faces was due to a voluntary strategy, [Experiment 4](#) employed a condition where the identity of each face switched every 500 ms during tracking. The results showed impaired performances compared to the standard condition where the identity of each face remained the same during tracking. Because identity switching is likely to result in higher resource demands for identity processing, the tracking impairment may be caused by involuntary identity processing.

To examine how tracking performance is modulated by mandatory and effortful target face encoding, participants in [Experiment 5](#) either performed the tracking task without being required to learn target faces or performed both the tracking and identity recognition tasks. We found that the tracking performance was more impaired when participants explicitly intended to learn the target faces, suggesting that conscious effort to learn target faces can result in a tradeoff between tracking and identity recognition.

These experiments have collectively demonstrated that representations are to some extent content addressable in multiple-face tracking. This is consistent with Horowitz et al.'s (2007) recent finding in object tracking. Our experiments offer clear evidence that the identity of target faces was processed to some level, whether or not it was task irrelevant. It lends support to the object file theory (Kahneman et al., 1992), which assumes identity encoding within a capacity limit and allows for content addressable information. However, the most intriguing finding in this study is that multiple-face tracking is not simply content addressable but also involves certain level of mandatory processing of facial identity. Our discussion will mainly focus on this issue and the role of attention in target processing and tracking performance.

Mandatory identity processing in multiple-face tracking

One of our main findings is that tracking performance can be impaired if the faces are upright and different from each other (see [Experiment 2](#)). This implies a degree of mandatory identity processing because encoding facial identity was task irrelevant. The finding contrast sharply with some of prior studies. For example, Scholl et al. (1999) suggest that unlike spatiotemporal properties, featural properties may not be encoded in MOT. Klieger et al. (2004) found that individuating objects by color in MOT did not change tracking performance, indicating a lack of feature or identity processing. Based on this kind of evidence, the FINST theory contends that the early vision responsible for MOT is “feature-blind” (Pylyshyn, 2004). Although some recent studies found that certain object features do get registered and play a role in MOT (Horowitz et al., 2007; Oksama & Hyönä, 2004, 2008), no evidence has suggested that identity processing in MOT is mandatory. Perhaps the most striking difference between our finding and prior literature is that while different faces impaired tracking performance in our study, different objects could facilitate tracking performance (Horowitz et al., 2007).

It is also worth noting that some explicit feature processing tasks in MOT do not affect tracking performance. In a dual task, where participants tracked multiple objects and monitored the object color change, tracking performance was not impaired, provided the monitoring response was made at the end of the trial rather than during tracking (Leonard & Pylyshyn, 2003). This was not true in our study, where tracking performance was impaired even though the identity recognition task was performed at the end of the trial ([Experiment 5](#)). This may demonstrate that features of objects and faces are processed differently. Complex facial features such as eyebrows and pigments are known to have strong influence on face recognition (Sinha, Balas, Ostrovsky, & Russell, 2006).

Facial features are often processed holistically, whereas object features are more likely to be processed in a piecemeal fashion.

These discrepancies may reflect fundamental differences between face and object processing in multiple target tracking. Perhaps the visual system is predisposed to discriminate faces and to encode facial identities at the expense of other visual information. This could underlie the deficit of tracking unique upright faces. When the same faces are inverted, the abandonment of configural processing and face discrimination releases attentional resources for the tracking, hence the better tracking performance. The distinctions between objects, on the other hand, are often less important for the visual system. Thus the advantage of tracking unique objects could be due to quite different reasons. As Horowitz et al. (2007) point out, if a target and a distractor are identical objects in MOT, they can be swapped or confused more easily by accident especially when they approach each other too closely. In addition, observers could search and recover a lost target by its distinct features. However, due to unmatched complexity of stimuli and experimental procedure between our study and others, conclusive evidence for the difference between face and object tracking will have to await more rigorous future examinations. Research on neural substrates of location–identity binding should also help resolve this issue. It is well known that the dorsal and the ventral streams of the visual pathways are specialized in processing the “where” and the “what” information, respectively. It would be interesting to know how these systems react to where–what binding differently when faces and objects are involved in the tracking tasks.

Attention and processing of facial identity

Experiments in this study have shown that in addition to some level of mandatory processing, attentional resources can also be consciously allocated for location tracking and identity encoding. Due to the nature of experimental manipulation, maximum resources in this study were either recruited for identity processing where only face learning was explicitly required (Experiment 1) or for location tracking where only tracking was explicitly required (Experiment 2). Attentional resources could also be more balanced between these two tasks (Experiments 3 and 4). We assessed the cost of identity processing for tracking by comparing tracking performance in these manipulations (Experiment 5).

Contrary to the view that processing of facial identity requires no attention, the results from Experiments 2 and 5 showed that identity processing interferes with tracking performance. The interference effect was likely to be caused by a divided attention to processing of facial identity. The results in Experiments 3 and 5 showed that target faces are more likely to be selected for identity processing whereas nontargets may be poorly processed or

inhibited during tracking. This presents further evidence against the assumption that face stimuli simply trigger identity processing. It shows that only selected faces receive identity processing.

The interference effect also creates difficulty for the view that facial identity is processed by a separate attentional mechanism (Palermo & Rhodes, 2002). If separate resources are available for processing facial identity, learning faces should not interfere with tracking. Quite to the contrary, the results in our experiments showed that attention for processing of facial identity is shared with the tracking task.

The tradeoff between identity encoding and tracking revealed in our experiments demonstrates a resource sharing where the same attentional system is used both for tracking and for face identification. It casts doubt on the view that the tracking aspect of MOT is automatic and nonattentional. If this were the case, target tracking should not have been detrimental to identity processing in Experiment 1. The finding in our study is consistent with Jackson and Raymond (2006) who also demonstrated resource sharing for face identification. Since faces may be just the type of stimuli that preferentially capture attention and produce certain level of mandatory processing, some of the attentional resources for tracking could be taken away, resulting in reduced tracking performance as shown in Experiments 2, 4, and 5. Since identity encoding is based on successful tracking, poor tracking performance could lead to impaired performance for face recognition in return. Tracking would directly influence identity processing because it creates a competition for limited attentional resources. This could explain why recognition performance was better after the faces had been learned in the static rather than the tracking condition in Experiment 1. If the tracking task is easy and demands less attention, the interference effect could become negligible, as was the case in Experiment 2 where the tracking performance for different faces and identical faces was indistinguishable when only three targets were used in for tracking.

When two tasks depend on a common set of resources, there is a potential for bidirectional interference, where each task interferes with the performance of the other, producing a joint “concurrency cost” (Brown, 2006; Navon & Gopher, 1979). Results in Experiments 1 and 2 showed bidirectional interference between identity processing and tracking. Some researchers make a distinction between visual attention and more general central executive attention (Johnston, McCann, & Remington, 1995). Tombu and Seiffert (2008) argue that tracking requires central attention. Information processing can be carried out in parallel, but it is subject to a central capacity. Consistent with this theory, Alvarez, Horowitz, Arsenio, DiMase, and Wolfe (2005) showed that both the tracking and auditory tasks in their study relied upon common attentional resources. The results of the bidirectional interference between face processing and tracking found

in this study also supports this interpretation. The common central resources may consist of storage components and control processes in the working memory. The control processes for fixations, pursuit eye movements, and the spatial extent of attentional focus are likely to be used for both tracking and identity processing task. However, identity processing may require longer and more narrowly focused attention on specific features and configural information of each face whereas tracking may require more transient and spatially extended attention to compute spatiotemporal continuity of multiple moving targets. The control processes may be used alternately for the two tasks, whereas the emphasis on each task may be partially modulated by task demands.

Finally, results in our study demonstrate that the attentional system allows encoding of multiple identities in the tracking task. Although the attentional limit for identity processing may still be just one face at a time (Bindemann et al., 2005), a quick target switching serial model of attention (e.g., Oksama & Hyönä, 2008) may account for encoding of multiple faces over time.

Conclusion

The MOT paradigm is now a principal tool for studying the connection between attention and object perception. However, the issue of dynamic location–identity binding that is so vital in social interaction has not yet been given enough attention. In this study, we sought to bridge this gap by providing empirical data on identity processing in multiple-face tracking. Our data demonstrate that target identity is not only processed when it is a task requirement, it can also be processed when it is task irrelevant. Unlike object tracking where performance can either remain the same or even improve when unique rather than identical objects are used, our data showed that tracking upright unique faces can impair tracking performance relative to tracking upright identical faces or inverted unique faces. This may be taken as evidence for some mandatory processing of facial identity and for a fundamental difference between dynamic identity–location binding for faces and objects. Our data also show that tracking and identity processing share the same attentional resources. The observer can consciously manipulate and allocate the resources to some level according to the task demands.

Acknowledgments

This research was supported in part by grants from 973 Program (2006CB303101), the National Natural Science

Foundation of China (90820305, 30500157, 30600182), the Royal Society China–UK Science Networks, and K. C. Wong Education Foundation. We thank Dr. Shan Shiguang for providing the face database and our reviewers for their many constructive comments and suggestions.

Commercial relationships: none.

Corresponding authors: Wenfeng Chen, Chang Hong Liu, Xiaolan Fu.

Email: chenwf@psych.ac.cn; c.h.liu@hull.ac.uk; fuxl@psych.ac.cn.

Address: 4A Datun Road, Chaoyang District, Beijing 100101, China; Cottingham Road, Hull HU6 7RX, UK.

References

- Alvarez, G., Horowitz, T., Arsenio, H., DiMase, J., & Wolfe, J. (2005). Do multielement visual tracking and visual search draw continuously on the same visual attention resources? *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 643–667. [[PubMed](#)]
- Awh, E., Serences, J., Laurey, P., Dhaliwal, H., van der Jagt, T., & Dassonville, P. (2004). Evidence against a central bottleneck during the attentional blink: Multiple channels for configural and featural processing. *Cognitive Psychology*, *48*, 95–126. [[PubMed](#)]
- Bahrami, B. (2003). Object property encoding and change blindness in multiple object tracking. *Visual Cognition*, *10*, 949–963.
- Bindemann, M., Burton, A. M., & Jenkins, R. (2005). Capacity limits for face processing. *Cognition*, *98*, 177–197. [[PubMed](#)]
- Boutet, I., Gentes-Hawn, A., & Chaudhuri, A. (2002). The influence of attention on holistic face encoding. *Cognition*, *84*, 321–341. [[PubMed](#)]
- Brown, S. W. (2006). Timing and executive function: Bidirectional interference between concurrent temporal production and randomization tasks. *Memory & Cognition*, *34*, 1464–1471. [[PubMed](#)] [[Article](#)]
- Brown, V., Huey, D., & Findlay, J. M. (1997). Face detection in peripheral vision: Do faces pop out? *Perception*, *26*, 1555–1570. [[PubMed](#)]
- Bruce, V., Hanna, E., Dench, N., Healey, P., & Burton, M. (1992). The importance of “mass” in line drawings of faces. *Applied Cognitive Psychology*, *6*, 619–628.
- Bruce, V., & Humphreys, G. W. (1994). Recognizing objects and faces. In V. Bruce & G. W. Humphreys (Eds.), *Object and face recognition* (pp. 141–180). Hove, England: Lawrence Erlbaum Associates.

- Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, *9*, 349–354. [PubMed]
- Downing, P. E., Liu, J., & Kanwisher, N. (2001). Testing cognitive models of visual attention with fMRI and MEG. *Neuropsychologia*, *39*, 1329–1342. [PubMed]
- Farah, M. J., Wilson, K. D., Drain, H. M., & Tanaka, J. R. (1998). What is “special” about face perception? *Psychological Review*, *105*, 482–498. [PubMed]
- Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X., et al. (2008). The CAS-PEAL large-scale Chinese face database and baseline evaluations. *IEEE Transactions on Systems, Man and Cybernetics*, *38*, 149–161. [Article]
- 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. [PubMed] [Article]
- Jackson, M., & Raymond, J. E. (2006). The role of attention and familiarity in face identification. *Perception & Psychophysics*, *68*, 543–557. [PubMed] [Article]
- Johnston, J., McCann, R., & Remington, R. (1995). Chronometric evidence for two types of attention. *Psychological Science*, *6*, 365–369.
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 175–219. [PubMed]
- Klieger, S. B., Horowitz, T. S., & Wolfe, J. M. (2004). Is multiple object tracking colorblind? [Abstract]. *Journal of Vision*, *4*(8):363, 363a, <http://journalofvision.org/4/8/363/>, doi:10.1167/4.8.363.
- Kuehn, S. M., & Jolicoeur, P. (1994). Impact of quality of the image, orientation, and similarity of the stimuli on visual search for faces. *Perception*, *23*, 95–122. [PubMed]
- Lavie, N., Ro, T., & Russell, C. (2003). The role of perceptual load in processing distractor faces. *Psychological Science*, *14*, 510–515. [PubMed]
- Leder, H. (1996). Line drawings of faces reduce configural processing. *Perception*, *25*, 355–366. [PubMed]
- Leonard, C., & Pylyshyn, Z. W. (2003). Measuring the attentional demand of multiple object tracking (MOT) [Abstract]. *Journal of Vision*, *3*(9):582, 582a, <http://journalofvision.org/3/9/582/>, doi:10.1167/3.9.582.
- Navon, D., & Gopher, D. (1979). On the economy of the human processing system. *Psychological Review*, *86*, 214–255.
- Nothdurft, H. C. (1993). Faces and facial expressions do not pop out. *Perception*, *22*, 1287–1298.
- 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.
- Oksama, L., & Hyönä, J. (2008). Dynamic binding of identity and location information: A serial model of multiple identity tracking. *Cognitive Psychology*, *56*, 237–283. [PubMed]
- Palermo, R., & Rhodes, G. (2002). The influence of divided attention on holistic face perception. *Cognition*, *82*, 225–257. [PubMed]
- Palermo, R., & Rhodes, G. (2007). Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia*, *45*, 75–92. [PubMed]
- Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, *32*, 65–97. [PubMed]
- Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, *80*, 127–158. [PubMed]
- Pylyshyn, Z. W. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual Cognition*, *11*, 801–822. [Article]
- Russell, R., Biederman, I., Nederhouser, M., & Sinha, P. (2007). The utility of surface reflectance for the recognition of upright and inverted faces. *Vision Research*, *47*, 157–165. [Article]
- Saiki, J. (2003a). Feature binding in object-file representations of multiple moving items. *Journal of Vision*, *3*(1):2, 6–21, <http://journalofvision.org/3/1/2/>, doi:10.1167/3.1.2. [PubMed] [Article]
- Saiki, J. (2003b). Spatiotemporal characteristics of dynamic feature binding in visual working memory. *Vision Research*, *43*, 2107–2123. [PubMed]
- Scholl, B. J., Pylyshyn, Z. W., & Kranconeri, S. L. (1999). When are featural and spatiotemporal properties encoded as a result of attentional allocation? *Investigative Ophthalmology and Visual Science*, *40*, 797.
- Searcy, J. H., & Bartlett, J. C. (1996). Inversion and processing of component and spatial-relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 904–915. [PubMed]
- Sinha, P., Balas, B. J., Ostrovsky, Y., & Russell, R. (2006). Face recognition by humans: 19 results all computer vision researchers should know about. *Proceedings of the IEEE*, *94*, 1948–1962. [Article]
- Tombu, M., & Seiffert, A. E. (2008). Attentional costs in multiple-object tracking. *Cognition*, *108*, 1–25. [PubMed]

- Vuong, Q. C., Peissig, J. J., Harrison, M. C., & Tarr, M. J. (2005). The role of surface pigmentation for recognition revealed by contrast reversal in faces and Greebles. *Vision Research*, *45*, 1213–1223. [[PubMed](#)]
- Wojciulik, E., Kanwisher, N., & Driver, J. (1998). Covert visual attention modulates face-specific activity in the human fusiform gyrus: fMRI study. *Journal of Neurophysiology*, *79*, 1574–1578. [[PubMed](#)] [[Article](#)]
- Young, A. W., Ellis, A. W., Flude, B. M., McWeeney, K. H., & Hay, D. C. (1986). Face–name interference. *Journal of Experimental Psychology: Human Perception and Performance*, *12*, 466–475. [[PubMed](#)]