Attention, Perception, & Psychophysics (2020) 82:1088–1098
https://doi.org/10.3758/s13414-019-01881-7

Interacting hands draw attention during scene observation

Ryosuke Niimi[1]

Published online: 14 November 2019

Abstract
In this study I examined the role of the hands in scene perception. In Experiment 1, eye movements during free observation of natural scenes were analyzed. Fixations to faces and hands were compared under several conditions, including scenes with and without faces, with and without hands, and without a person. The hands were either resting (e.g., lying on the knees) or interacting with objects (e.g., holding a bottle). Faces held an absolute attentional advantage, regardless of hand presence. Importantly, fixations to interacting hands were faster and more frequent than those to resting hands, suggesting attentional priority to interacting hands. The interacting-hand advantage could not be attributed to perceptual saliency or to the hand-owner (i.e., the depicted person) gaze being directed at the interacting hand. Experiment 2 confirmed the interacting-hand advantage in a visual search paradigm with more controlled stimuli. The present results indicate that the key to understanding the role of attention in person perception is the competitive interaction among objects such as faces, hands, and objects interacting with the person.

Keywords Eye movements · Face perception · Hand perception · Human body · Natural scenes · Visual search

It is well known that the face is critical to visual person perception. The human visual system is fine-tuned to process faces, which have an attentional advantage over other objects (Hershler & Hochstein, 2005; Langton, Law, Burton, & Schweinberger, 2008; but see also VanRullen, 2006). However, the role of nonfacial body parts in person perception is still unclear. Neuroscientific evidence suggests that humans and monkeys are equipped with neural mechanisms attuned to nonfacial body parts (Peelen & Downing, 2007). Functional brain imaging studies have demonstrated the existence of brain regions dedicated to body perception in both humans (Downing, Jiang, Shuman, & Kanwisher, 2001) and monkeys (Pinsk, Desimone, Moore, Gross, & Kastner, 2005). Hands in particular have been objects of research interest: "hand cells" as well as "face cells" have often been found in macaque monkey temporal lobes (Desimone, Albright, Gross, & Bruce, 1984). A hand-specific visual cortex is also likely to exist in the human brain (Bracci, Ietswaart, Peelen, & Cavina-Pratesi, 2010; Op de Beeck, Brants, Baeck, & Wagemans, 2010) and is likely to underlie not only hand detection, but the perception of hand action (Perini, Caramazza, & Peelen, 2014), which seems critical to person perception.

In contrast to these clear-cut findings, psychological evidence from behavioral experiments has often been controversial. How and to what extent are nonfacial body parts processed for person perception? A body without a face is sufficient for person identification (Rice, Phillips, Natu, An, & O'Toole, 2013; Robbins & Coltheart, 2012). However, bodies contribute much less to person recognition than faces (Burton, Wilson, Cowan, & Bruce, 1999) and can even have no effect when a face is present and/or when the stimuli are nonmoving still images (O'Toole et al., 2011; Robbins & Coltheart, 2012; Simhi & Yovel, 2016).

The effects of nonfacial body parts on attention are likewise unclear. Whole-body human figures (those including both the face and body) draw attention. When observers search a display showing multiple scene pictures, their attention is biased toward scenes with human figures (Fletcher-Watson, Findlay, Leekam, & Benson, 2008; Mayer, Vuong, & Thornton, 2015). Whole-body silhouettes are less subject to inattentional blindness than nonhuman objects (Downing, Bray, Rogers, & Childs, 2004). Less evidence is available for nonface body parts, however. Using a probe-dot detection task following object picture presentations, Morrisey and Rutherford (2013) reported that whole bodies, hands, and feet drew more attention than nonhuman objects. In contrast, hands and nonhuman objects were equally subject to inattentional blindness (Downing et al., 2004). Ro, Friggel, and Lavie (2007) suggested that both faces and other body parts have an attentional advantage over nonhuman objects in visual search. However,

✉ Ryosuke Niimi
niimi@human.niigata-u.ac.jp

[1] Faculty of Humanities, Niigata University, Niigata, Japan

they did not directly compare them. Bindemann, Scheepers, Ferguson, and Burton (2010) showed that the time needed for person detection in a natural scene was equivalent for face-only targets and headless body targets, which seems to contradict the idea that the face is special.

A key to understanding the role of nonfacial body parts in person perception is to consider them in relation to the face. Most of the attentional studies reviewed above presented body part stimuli separately, while faces and body parts were simultaneously present in everyday scenes. Eyetracking studies using natural scene stimuli have focused on this issue. Birmingham, Bischof, and Kingstone (2008b) recorded eye movements while participants freely observed scenes containing persons. They found that participants fixated more on eyes and heads than on bodies. Virtually identical results were obtained for performed tasks other than free-viewing (Birmingham, Bischof, & Kingstone, 2008a). These findings suggest an attentional priority for heads and faces over nonfacial body parts. Comparable findings were also reported by Fletcher-Watson and colleagues (Fletcher-Watson et al., 2008; Fletcher-Watson, Leekam, Benson, Frank, & Findlay, 2009).

However, no study has yet examined how hands and faces jointly guide attention during natural scene observation. Among the body parts, hands are of particular interest in understanding person perception and scene perception. Hands as well as faces are easily observed by others and are essential for understanding their owners' behaviors and intentions. Consistent with this view, some brain regions are more strongly activated by hands interacting with objects in functional ways (Johnson-Frey et al., 2003). We use both facial expressions and hand gestures for communication. Furthermore, face and hands in tandem may guide others' attention (e.g., gaze and pointing). These facts suggest the hypothesis that hands may play a special role in attention and visual perception compared with other body parts (e.g., torsos, feet). Kano and Tomonaga (2009) compared gaze times for faces, torsos, arms, and legs, but the arms received fewer fixations than faces, and no attentional bias to arms was found. Thus, it seems critical to focus on hands rather than on arms.

In Experiment 1, I investigated eye movements during scene observation and compared attention to faces with attention to hands. Possible interactions between hands and faces were of particular interest. Hand action and posture are also critical issues in understanding the role of hands in person perception. Thus, I compared fixations for resting hands (e.g., those just lying on the knees) to fixations for hands interacting with other objects (e.g., playing the piano). The observers viewed the scenes freely, and their eye movements were analyzed from various perspectives to explore what transpired during the observation of human figures in natural scenes.

Experiment 1

Method

Design Six scene conditions were used (Fig. 1a). Each condition contained 24 unique scenes. Each participant observed each scene once, resulting in 144 trials.

The *object* condition scene contained one salient, central object but did not contain a human figure. In the other conditions, each scene contained one human figure. No scene included two or more persons, with the exception of some scenes that contained small, blurry human figures in the background. In other words, no scenes showed social interactions. For the effects of the number of persons and their social interactions on scene perception, see Birmingham et al. (2008a, 2008b).

The *face* condition scene contained a face but not hands. There were two *hand* (without a face) conditions, one for resting hands and another for interacting hands. The resting hands either did not touch anything or were just placed on the owners' bodies (e.g., on the knee). The interacting hands touched or manipulated objects (e.g., a pen, handle, or bottle) in functional ways, suggesting their owners' interactions with the objects. Hands with gestures (e.g., pointing or waving) were avoided. These four conditions, in turn, each included a single area of interest (AOI) (an object, face, or hand).

The remaining two conditions were two-AOI conditions, in which each scene contained both a face and a hand. In one of these conditions, the scenes contained a face and a resting hand, and in the other they contained a face and an interacting hand.

Thus, in total there were eight categories of AOIs: object, face (without hand), resting hand (without face), interacting hand (without face), face (with resting hand), face (with interacting hand), resting hand (with face), and interacting hand (with face). No scenes had two face AOIs or two hand AOIs. The primary purpose of this experiment was to examine how the speed and number of fixations would vary among these eight categories of AOI (Fig. 1a).

Participants Seventeen individuals were paid for their participation. They all were graduate or undergraduate students. The results from four of them were not reported in this article because of strabismus (one participant) and a relatively low proportion of valid gaze data (see the Results). Of the remaining 13 participants, ten were male and three female, 19–31 years of age (mean = 22.5). They all reported normal or corrected-to-normal vision. Seven of the participants wore eyeglasses during the experiment, whereas six had naked eyes. Since contact lenses often yield poor calibration in the system used here, I recruited only individuals with eyeglasses or naked eyes. Their written informed consent was obtained in advance.

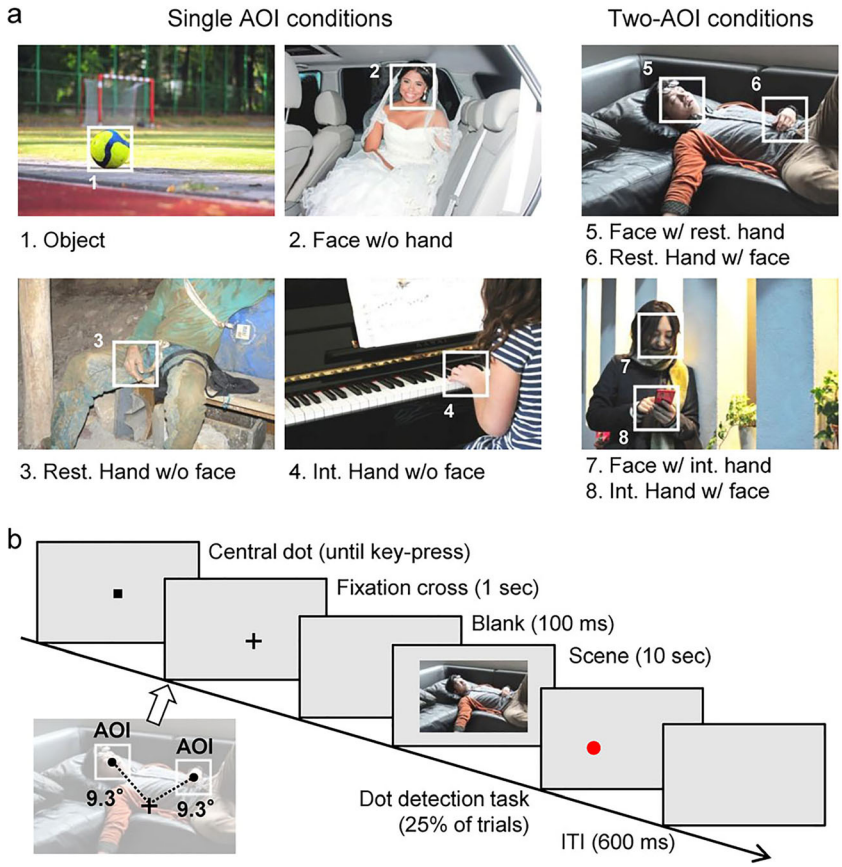

Fig. 1 **a** The six scene conditions and the eight areas of interest (AOIs). AOIs are indicated by white frames and numbers in this figure (not shown during the experiments). **b** Procedure for each trial. The position of the fixation cross was set permanently 9.3° away from the AOI center(s) in the following scene. Int. Hand, interacting hand. Rest. Hand, resting hand

Apparatus One eye (the self-reported dominant eye) was recorded while the participants observed the stimuli with both eyes. Each participant's dominant eye was illuminated by infrared (IR) LEDs (peak wavelength, 890 nm) and recorded by a 120-Hz CCD camera (OptiTrack V120 Slim, NaturalPoint, Inc.). The LEDs and the camera were placed in front of the stimulus screen and were fixed to the desk. GazeParser/SimpleGazeTracker (Sogo, 2013), an open-source software for video-based eye tracking, detected the pupil and Purkinje reflection and estimated the gaze position. The eyetracking system was assembled by the author and controlled by a personal computer (Dell OptiPlex 760) and SGT Toolbox software (http://sgttoolbox.sourceforge.net/).

The experiment was controlled using a personal computer (Dell Precision T3400) with an Ubuntu 12.04 operating system. A computer script for GNU Octave 3.2.4 with the SGT Toolbox 0.2.2 and Psychophysics Toolbox 3.0.11 (Kleiner, Brainard, & Pelli, 2007) controlled the eyetracking system, presented the stimuli, and recorded participants' key-press responses.

The experiment was conducted in a normally illuminated room. Stimuli were presented on a 24-in. LCD screen (Dell U2410) driven by an AMD FirePro V3900 GPU. The screen was set to 60 Hz and had a 1,920 × 1,200-pixel resolution. Participant viewpoint was fixed at 57 cm from the screen by a chin rest.

Stimuli The scene stimuli were 144 color images obtained from photo stock packages and services. They were 31.8 × 21.2° (1,200 × 800 pixels) in size and presented in the center of the screen. The background of the display was uniformly gray.

Each scene contained one or two AOIs (see the Design section and Fig. 1a). The AOIs were defined as square regions enclosing the target object/face/hand. Every AOI subtended 5.4 × 5.4° of visual angle. I selected the scenes to control for the spatial distribution of the AOIs and to ensure that the averages and standard deviations of the distances from the center of the scene picture to the AOI center were roughly equal for all eight categories of AOI.

Task and procedure The participants' task was free viewing. They observed the scenes as they wished. Each participant observed 144 scenes (144 trials), with self-paced breaks after every 24 trials. In addition, the participants were asked to make a quick key-press response to a red dot that appeared after scene presentation in 25% of the trials (Fig. 1b). This dot

detection task was introduced in order to maintain participants' arousal.

The experiment started with calibration of the eyetracking system. The camera and infrared illumination were trained on the self-reported dominant eye. The calibration procedure for SimpleGazeTracker was run to record the positions of the pupil and the Purkinje reflection for nine points on the screen. If the calibration failed, the other eye was tested.

Each trial began with a central black dot (Fig. 1b), which was presented until the participant pressed the "5" key on the numeric keypad to start the trial. A black fixation cross was then shown for 1 s. The participants were instructed to fixate on this cross. The position of the fixation cross was determined in advance to be 9.3° from any AOI, to ensure that the gaze positions at the time of scene onset were equidistant from any AOI. For single-AOI scenes, the fixation cross was positioned on the line through the AOI center and the scene center. Since two points on this line were 9.3° from the AOI center, the point closer to the scene center was adopted. For two-AOI scenes, two points were 9.3° from the two AOI centers, and the point closer to the screen center was adopted (see Fig. 1b inset).

Following the fixation cross, a blank gray screen appeared for 100 ms. A scene picture was then presented at the center of the screen for 10 s. The participants observed the scene freely. After the scene, a blank display was shown during a 0.6-s intertrial interval (ITI). The next trial then followed.

In 36 trials (25%), a dot detection task was introduced after scene presentation (Fig. 1b). The trials with the dot detection task were randomly chosen under the constraint of six trials per scene condition. A red dot was presented in a random position within the area where a scene was shown. The participants were asked to press the "8" key of the numeric keypad as fast as possible when the red dot appeared. When the "8" key was pressed or 5 s had passed, the screen turned blank, followed by a 0.6-s ITI. The next trial then started.

After 72 trials (i.e., halfway), the participants were allowed to release their heads from the chin rest. After a self-paced break, the eyetracking system was recalibrated and the experiment was resumed. The experiment took approximately 1 h.

Results

Gaze positions during scene presentation (10 s, 1,200 samples) were analyzed. The SGT Toolbox software generated the estimated gaze position in terms of coordinates on the screen, and these data were analyzed by custom-made computer code running on GNU Octave.

Since gaze position could not be measured when the eyelid covered the pupil or the participant directed his/her gaze outside of the screen, a proportion of the valid samples of gaze position were assessed for every trial. Trials with 50% or fewer valid samples were discarded as invalid. Because three participants had 10% or more invalid trials, their results were excluded from the analysis. The remaining 13 participants yielded 0.7% invalid trials on average. Hence, 1,858 valid trials were analyzed. The mean proportion of valid samples from these trials was 92.1%.

Fixations were detected as 50-ms or longer successions of gaze shifts below 0.5° across two consecutive samples. Fixation position was defined as the averaged gaze positions of a single fixation. An average of 23.1 fixations emerged per trial. An analysis of variance (ANOVA), with one within-participant factor of scene condition (Fig. 1a), indicated that the number of fixations was dependent on the scene condition [$F(5, 60) = 2.74$, $p = .027$, $\eta_p^2 = .19$], although multiple comparisons (Ryan's method, $\alpha = .05$) did not yield any significantly different pairs.

Fixation time within AOI In each trial, I measured the sum of the durations of the fixations for each AOI. As is shown in Fig. 2a, the mean fixation time was longer for the object and face

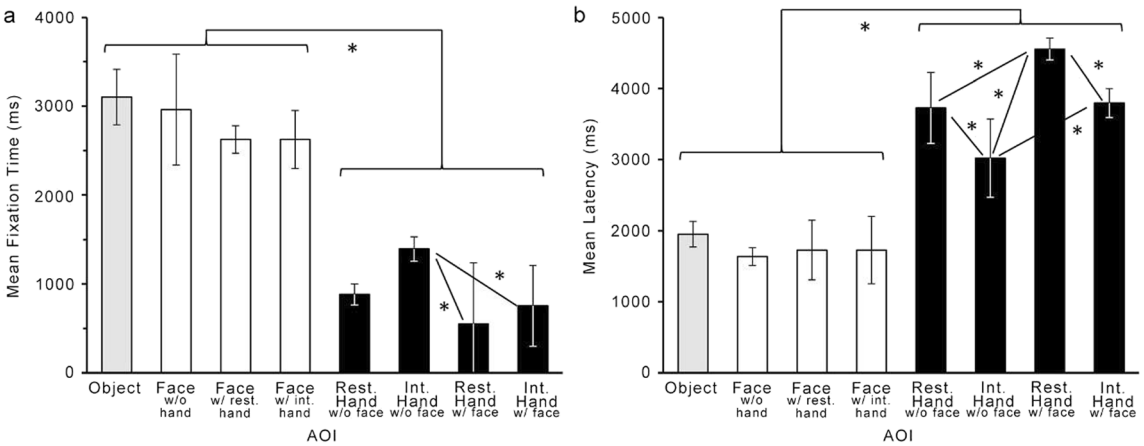

Fig. 2 Mean total fixation times for each area of interest (AOI) (**a**) and mean latencies of first fixations for each AOI (**b**). Error bars indicate a 95% confidence interval (CI) of the mean. $^*p < .05$

AOIs than for the hand AOIs. An ANOVA with one within-participant factor of AOI confirmed that fixation time varied among the eight AOIs shown in Fig. 1a [$F(7, 84) = 44.7$, $p < .001$, $\eta_p^2 = .79$]. Multiple comparisons (Ryan's method, $\alpha = .05$) revealed that all four hand AOIs yielded significantly shorter fixation times than did the other four AOIs, indicating less fixation on the hands than on faces and salient objects. Multiple comparisons also revealed that interacting hands without faces received significantly longer fixations than did the two hand AOIs with faces.

This pattern showed the absolute attentional advantage of faces over hands. Faces were fixated on longer than hands, and faces with and without hands received equal lengths of fixation.

Fixation latency for AOI For each AOI of each trial, fixation latency was measured from scene onset to the beginning of the first fixation for that AOI. The fixation latency was much shorter for object and face AOIs than for hand AOIs (Fig. 2b). A one-factor ANOVA confirmed variation in fixation latency among the eight AOIs [$F(7, 84) = 63.8$, $p < .001$, $\eta_p^2 = .84$]. Multiple comparisons (Ryan's method, $\alpha = .05$) revealed that the latencies of the four hand AOIs were significantly longer than those of the other AOIs. Furthermore, among the four hand AOIs, interacting-hand AOIs yielded significantly shorter latencies than did resting-hand AOIs. I also found that the presence of a face increased the latency to the hands. Resting hands without faces yielded significantly shorter latencies than resting hands with faces, and interacting hands without faces yielded significantly shorter latencies than interacting hands with faces. This pattern was virtually identical to that for fixation times: Faces had an absolute advantage over hands, and interacting hands had an advantage over resting hands.

One might assume that perceptual saliency was responsible for the interacting-hand advantage. Interacting hands might be more salient than resting hands because they take more complex forms or often hold salient objects (e.g., tools). This was not the case, however. Using a computer code provided by Harel (2012), perceptual saliency (Itti, Koch, & Niebur, 1998) was computed for every AOI. Each AOI saliency was divided by the average saliency of all pixels of the scene image, which yielded normalized saliency. One-factor ANOVA revealed that the normalized saliency varied significantly among the eight categories of AOI [$F(7, 184) = 3.63$, $p = .001$, $\eta_p^2 = .12$]. Multiple comparisons (Ryan's method, $\alpha = .05$) showed that object AOIs showed significantly higher saliency than the others. No reliable difference, however, emerged among the other AOI categories. There was no evidence that the interacting-hand AOIs were perceptually more salient than the resting-hand AOIs. Thus, the interacting-hand advantage could not be attributed to perceptual saliency.

Analysis of fixation sequences To examine the effects of faces and hands on the time course of the gaze shift, I analyzed the proportions of fixations within each AOI as a function of the sequential order of fixations. Figure 3 shows the proportions of trials in which each AOI was fixated on, by order of fixation after scene onset. Because the proportions of face and hand fixations are dependent on each other in two-AOI scenes, I analyzed the data for single-AOI scenes and two-AOI scenes separately.

For the single-AOI scenes (Fig. 3a), faces and objects received more first fixations than the hand AOIs. Critically, more first fixations were found on interacting than on resting hands. An ANOVA (4 AOIs × Ordinal Fixation [1–15]) revealed significant main effects of AOI [$F(3, 36) = 83.0$, $p < .001$, $\eta_p^2 = .87$], ordinal fixation [$F(14, 168) = 38.3$, $p < .001$, $\eta_p^2 = .76$], and their interaction [$F(42, 504) = 12.5$, $p < .001$,

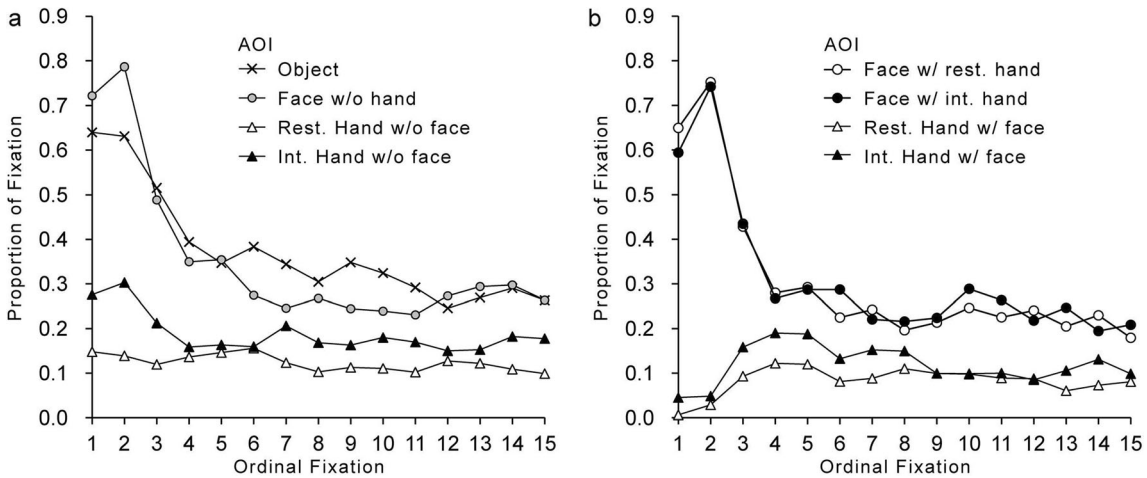

Fig. 3 Proportions of fixated areas of interest (AOIs) for each fixation: Results for (**a**) single-AOI scenes and (**b**) two-AOI scenes

η_p^2 = .51]. Overall, faces and objects received more fixations than hands. Multiple comparisons on the simple main effect of AOI (Ryan's method, α = .05) revealed that faces received more fixations than objects in the first and second fixations, and interacting hands received more fixations than resting hands from the first through the third fixations. These results signify attentional advantages for faces and interacting hands in initial fixations in natural scenes. Interestingly, object AOIs received significantly more fixations than facial AOIs in the later fixation sequences (sixth, seventh, ninth, and tenth). This pattern suggests that faces summon attention at the initial stages of scene observation, but attention is later drawn to nonfacial regions of the scene.

The prominent advantage of faces also emerged for the two-AOI scenes (Fig. 3b): Faces received more fixations than hands for all ordinal fixations. However, the pattern of early fixations suggests a trade-off between faces and hands. Faces received the most fixations in the initial stages of viewing (first to third fixations), whereas fixations on hands increased later (third to fifth fixations). Another critical finding was that interacting hands received more fixations than resting hands. An ANOVA (AOI [resting/interacting hand] × Ordinal Fixation [1–15]) revealed significant main effects of AOI [$F(1, 12) = 15.9$, $p < .001$, $\eta_p^2 = .57$] and ordinal fixation [$F(14, 168) = 6.02$, $p < .001$, $\eta_p^2 = .33$], without an interaction effect. This finding is consistent with the shorter latency and longer fixation time for interacting-hand AOIs than for resting-hand AOIs (Fig. 2).

Effect of the depicted person's gaze in the scenes: Gaze at one's hand It is known that another person's direction of gaze influences observers' attentional orientation. Gaze direction of a face image acts as an endogenous attentional cue (Driver et al., 1999). This is also the case for complex naturalistic scenes (Dukewich, Klein, & Christie, 2008; Zwicker & Võ, 2010). It seems plausible that participants first fixated on faces and then shifted their attention to the objects gazed upon by the face in the scene. Moreover, it could be assumed that the interacting-hand advantage was also due to such attentional guidance because one is more likely to look at his or her own hand if it is interacting with something than when it is resting. However, this hypothesis cannot fully account for the interacting-hand advantage, which held even in scenes without faces (see Fig. 2 for the results from the interacting hand without a face AOI and the resting hand without a face AOI). An interacting hand itself is likely to yield attentional advantage over a resting hand.

It is still possible, however, that attentional guidance by gaze partly contributed to the interacting-hand advantage in the two-AOI scenes. To test whether hands gazed at by their owners received more fixations than hands not being gazed at, I divided the 24 scenes with faces and interacting hands into two groups: six scenes in which the interacting hands received their owner's gaze, and 18 scenes in which the interacting hands did not receive a gaze. The mean latency for the hand AOIs was 2,812.5 ms for the former (i.e., hands with a gaze) and 2,955.4 ms for the latter (hands without a gaze), which was not a statistically significant difference [$t(12) = 0.72$, p = .48, d = 0.21]. The mean fixation time was 923.5 ms for hands with a gaze and 704.7 ms for hands without a gaze. Again, this was not a significant difference [$t(12) = 1.75$, p = .11, d = 0.50]. However, an analysis of ordinal fixations revealed that hands with a gaze in the picture were more fixated on by participants than the hands without a gaze (Fig. 4). A two-way ANOVA (Hand Owner's Gaze [on/off hand] × Ordinal Fixation [1–15]) revealed a significant two-way interaction [$F(14, 168) = 2.11$, $p < .014$, $\eta_p^2 = .15$]. The simple main effect of the hand-owner's gaze was significant in the third, fourth, fifth, and seventh fixations (ps < .05), suggesting that a person's gaze direction guided the participants' attention toward that person's hands at a relatively early stage of scene perception, probably just after the initial fixation to the face (the first and second fixations).

Effect of the depicted person's gaze in the scenes: Gaze at the observer Another issue tested was the effects of gazes directed at the observer (i.e., the camera). Studies have demonstrated that mutual gaze, or eye contact, modifies recognition memory (Vuilleumier, George, Lister, Armony, & Driver, 2005) and hand movement (Wang, Newport, & Hamilton, 2011). I examined whether faces gazing toward the observer (mutual-gaze faces) received more fixation than faces gazing at other objects (averted-gaze faces). I classified the 24 scenes with faces (but without hands) into 10 scenes with mutual-gaze faces and 14 with averted-gaze faces. No difference in mean latencies for the face AOI was found between the groups

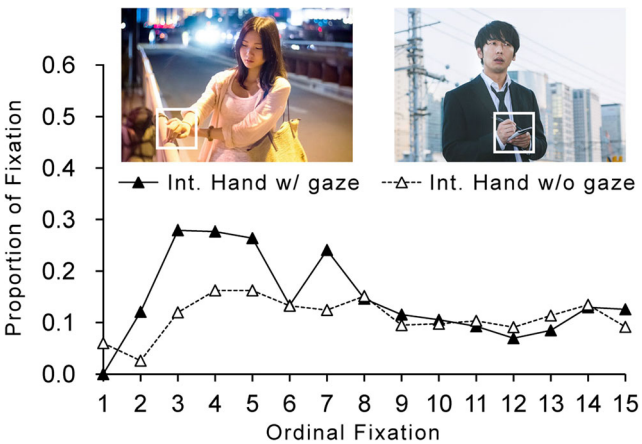

Fig. 4 Scenes with faces and interacting hands were divided into scenes in which the person in the scene was looking at his or her hand (Int. Hand w/ Gaze) and scenes in which the person was not looking at his or her hand (Int. Hand w/o Gaze). The proportions of fixations on the hand area of interest are shown

(1,527.9 ms for mutual-gaze faces and 1,652.5 ms for averted-gaze faces), $t(12) = 1.74$, $p = .11$, $d = 0.50$. In contrast, the mean fixation time was significantly longer for mutual-gaze faces (3,151.4 ms) than for averted-gaze faces (2,825.2 ms), $t(12) = 2.74$, $p = .018$, $d = 0.79$. In addition, the total scan path length during scene presentation (10 s) was shorter for scenes with mutual-gaze faces (mean = 14,980 pixels) than for scenes with averted-gaze faces (mean = 15,353 pixels), $t(12) = 3.00$, $p = .012$, $d = 0.86$. These results suggest that mutual-gaze faces held observers' attention longer than averted-gaze faces, but they did not draw observers' attention more quickly than averted-gaze faces.

Generalized linear mixed model (GLMM) analysis of fixations to hand Multiple factors appeared to affect gaze behavior toward the hands. To examine these factors in a unified framework, GLMM was applied to the latency data for hand AOI (the Gamma distribution was hypothesized, with log link function). I analyzed the 426 valid trials of the two-AOI scenes. Since the present findings suggested that the participants first fixated on faces and then shifted their gaze to hands, face AOI latency and face-to-hand distance (in pixels) were adopted as fixed-effect independent variables. The perceptual saliency (normalized by average image saliency) of hand AOI, hand action (interacting/resting), hand-owner's gaze on hand (on or away from the hand), and gaze to observer (mutual or averted) were incorporated into the model as well. The variables were standardized in advance. The random effect was the participants.

The analysis revealed that face–hand distance, hand action, and hand-owner's gaze on hand showed reliable effects on the latency to hand fixation (Table 1). Hand AOI saliency and gaze to observer yielded no reliable effect, and face latency showed little effect. These three variables had virtually no effect on the latency to hand fixation, as the Akaike information criterion changed only slightly after eliminating them from the model. The same analysis was conducted for fixation time data for the hand AOIs as well, which replicated the significant effects of face-hand distance and hand action. The hand-owner's gaze on hand did not account for the fixation time data, however.

These results further confirmed the following conclusions: (1) Interacting hands draw attention, and this effect cannot be accounted for by perceptual saliency; (2) a hand-owner's gaze guides observers' attention; and (3) hands are fixated on after faces.

Table 1 Results for fixed effects from the generalized linear mixed model for hand area-of-interest (AOI) latency data

	Estimate	(*SE*)	*p*	
Intercept	– 1.09	(0.16)	< .001	***
Slope				
Face AOI latency	0.87	(0.46)	.056	†
Face–hand distance	0.70	(0.27)	.009	**
Hand AOI saliency	– 0.04	(0.03)	.209	
Hand action (int./rest.)	– 0.20	(0.05)	< .001	***
Gaze on hand	– 0.20	(0.06)	.002	**
Gaze on observer	0.03	(0.05)	.574	

int., interacting hand; rest., resting hand. † $p < .1$, ** $p < .01$, *** $p < .001$

Discussion

Consistent with the findings of previous studies, the attentional advantage for faces was very prominent. Faces were attended to earlier and for longer periods than hands (Fig. 2). For scenes that included both faces and hands (the two-AOI conditions), most of the first fixations involved faces, and virtually none involved hands (Fig. 3b). As a consequence, hands received more fixation by around the third fixation than during the first. Thus, during the free observations of natural scenes, attentional priority to hands lags behind that to faces.

The critical finding was that interacting hands drew longer and quicker fixations than resting hands. The interacting-hand advantage was found in both the single-AOI and two-AOI conditions. Although the advantage may be partly attributable to the higher chance of the hand being looked at by the person in the scene (Fig. 4), it was still found in scenes without faces. Indeed, interacting hands drew more first fixations than resting hands if a face was absent (Fig. 3a). Hence, the interacting hand itself was likely the focus of attention.

However, the interacting-hand advantage might be attributable to more objects in the hand AOIs. By definition, the interacting hands were always in contact with other objects (e.g., smartphones). Although the perceptual saliency of interacting-hand AOIs did not differ from that of resting-hand AOIs, it was still possible that the presence of other salient object in interacting-hand AOIs might have yielded these results. Experiment 2 examined this issue with more controlled stimuli.

Experiment 2

In this experiment, using a visual search paradigm, I tested the effect of interacting hands on spatial attention in more controlled situation. The participants were asked to search for a predetermined target (hand, flower, or cup), in which the hand images were manipulated (interacting/resting). In the same way, nonhand target objects (flower and cup) were shown in interacting/noninteracting conditions; for instance, a flower partially occluded a door knob in the interacting condition. I tested whether the effect of interaction would appear in the visual search for hand, flower, or cup. It was hypothesized that an interacting hand target would be found faster than a

noninteracting (i.e., resting) hand target, whereas the search times for interacting flowers and cups would not differ from those for noninteracting flowers and cups.

Method

Stimuli In each trial, a visual search array of eight objects was presented (Fig. 5). Each array was generated by combining four component images (one image per quadrant). Each component image included two objects. There were four categories of component images: hand component images in which one hand (either interacting or noninteracting; i.e., resting) and one filler object were shown, flower component images (flower and filler object), cup component images (cup and filler object), and filler component images (two filler objects). Each search array consisted of these four component images—namely, any array included one hand, one flower, one cup, and five filler objects. The filler objects were chosen from ten everyday objects (comb, computer mouse, door knob, notebook, pen, remote control, ruler, scissors, smartphone, and tennis ball). All object images were achromatic and shown on a uniformly gray background. Search arrays were shown at the center of computer screen, subtending 1,200 (width) × 800 (height) in pixels, 30.2° × 20.4° in visual angle.

To generate search arrays, 70 component images were prepared in advance. For the interacting hand condition, ten component images of hands interacting with one filler object were made (e.g., a hand holding a door knob). For the noninteracting hand condition, ten images of a hand and filler object were made so that the hand and the object were detached. In the same manner, ten component images of an interacting-flower with a filler object were made so that the flower partially occluded the filler object. Note, however, that the "interacting" flower did not interact with the filler object in a functional way, but just occluded the filler object (cf. Green & Hummel, 2006; Kim, Biederman, & Juan, 2011). Then ten component images of noninteracting flowers with a filler object were made (the flower was detached from the filler object). Such 10 + 10 component images were similarly constructed for cups. If the attentional priority of an interacting hand observed in Experiment 1 was due to more objects flanking or in contact with the hand, the interacting targets would be found faster than noninteracting targets irrespective of target category (hand, flower, or cup). In addition, ten filler component images (five for interacting fillers and five for noninteracting fillers) were made.

As a result, one hand, one flower, and one cup were always shown in separate quadrants. Each search array contained two interacting component images and two noninteracting component images (see Fig. 5). The positions of the component images were randomized in every trial.

Participants Fifteen graduate or undergraduate students were paid for their participation. They all reported normal or corrected-to-normal vision. Their written informed consent was obtained in advance. No individual had participated in Experiment 1.

Apparatus Stimuli were presented on a 24-in. LCD screen with 1,920 × 1,200 pixel resolution. The experiment was controlled by computer script for Psychophysics Toolbox 3.0.11 (Kleiner et al., 2007) running on a personal computer (Dell Precision T3400). Participants' viewing distance was approximately 60 cm. Responses were obtained by a standard computer keyboard.

Task and procedure In each trial, a visual search array of eight objects was presented (see Stimuli section). The task was to find a category-defined target object (either hand, flower, or cup) and to report whether the target was on the left or right side of the array as quickly as possible. Participants pressed the F (J) key of the computer keyboard with the index finger of their left (right) hand if the target was found on the left (right) side.

Each participant conducted three blocks, a hand block in which they were required to search for a hand, a flower block (the search target was a flower), and a cup block (the search target was a cup). The order of the three blocks randomly

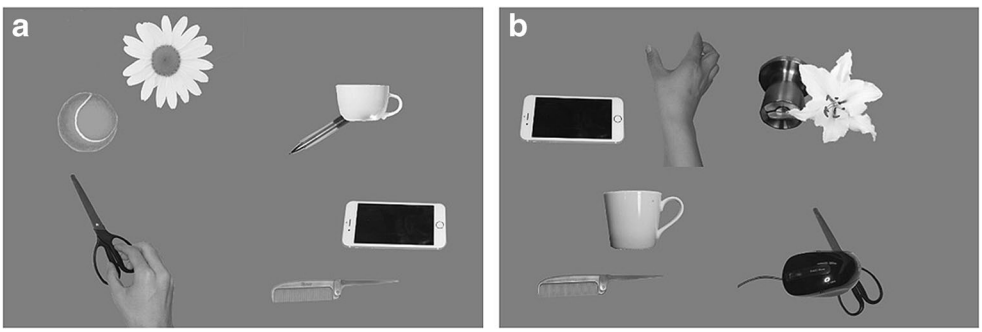

Fig. 5 Example search arrays in Experiment 2. Each array consisted of a hand, a flower, a cup, and five filler objects. **a** An array including an interacting hand, a noninteracting flower, and an interacting cup. **b** An array including a noninteracting hand, an interacting flower, and a noninteracting cup

varied among the participants. Each block consisted of 80 trials. The target object appeared on the left side in 40 trials and on the right side in 40 trials. An interacting target object appeared in 40 trials and a noninteracting target in 40 trials. The order of trials was randomized. As is described in the Stimuli section, every search array contained one hand, one flower, and one cup in any trial of any block. Identical sets of component images was used for the three blocks. Before each block, 20 practice trials were conducted.

Each trial started with a central fixation cross (black), which was presented until the participant pressed the space key. If the space key was pressed, the fixation cross turned white and disappeared 200 ms later. A blank display (300 ms) followed, then a search array was presented. The array was shown until response was made. No feedback on response accuracy was given. Following 400-ms ITI (blank screen), the next trial started.

Results

Figure 6 shows the results. Mean reaction times for the six conditions (Search Target [hand/flower/cup] × Target Interaction [interacting/noninteracting]) were calculated for each participant. Error responses were excluded from the analysis. Exceptionally fast (below 100 ms) or slow (2,000 ms or longer) responses were treated as errors. Error rates were very low (see Fig. 6).

To test the hypothesis that an interacting target was found faster than a noninteracting target in the hand search but not in the flower or cup search, I conducted a multiple comparison

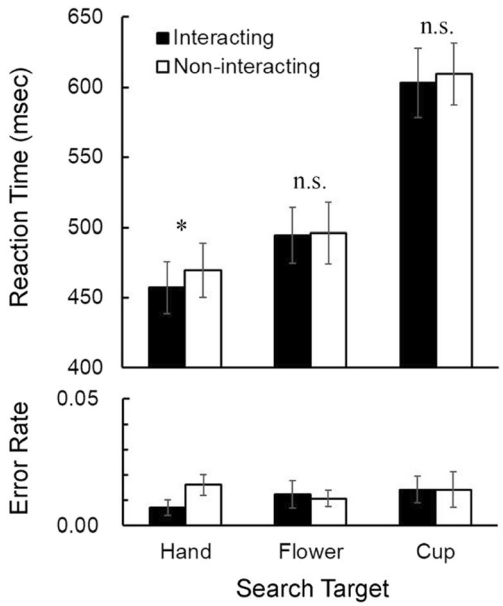

Fig. 6 Results of Experiment 2 (visual search): Mean reaction times of visual search for an interacting hand were shorter than search for a noninteracting hand. Error bars denote *SEM*s

(three within-participant *t* tests on the effect of target interaction) with the Bonferroni–Holm correction (α = .05). The result confirmed that interacting hands were found significantly faster than noninteracting hands [$t(14) = 3.14$, $p = .007$, $d = 0.84$], whereas such an effect of interaction was not significant for the flower search [$t(14) = 0.49$, $p = .632$, $d = 0.13$] or the cup search [$t(14) = 0.458$, $p = .654$, $d = 0.12$]. This pattern of results supported the hypothesis that an interacting hand had attentional priority over a noninteracting (i.e., resting) hand even when the number of objects in contact with the search target was controlled.

General discussion

The two experiments demonstrate that interacting hands have an attentional advantage in scene perception. Although the advantage was not as strong as that of faces, the results suggest that hands may play an important role in the visual perception of natural scenes with human figures.

It seems plausible that functional hand postures/actions are efficiently encoded in the visual system. Functional brain-imaging studies have demonstrated that some brain regions are sensitive to functional hand-body interactions (Bracci & Peelen, 2013; Johnson-Frey et al., 2003). Highly efficient visual processes in the neural network of those regions may contribute to the attentional advantage of interacting hands over resting hands.

The interacting-hand advantage may also reflect an adaptive attentional function dedicated to person perception. In the same vein, an attentional bias was found toward objects being gazed upon by a person in the scene (Driver et al., 1999; Zwicker & Võ, 2010). This too may facilitate person perception. Fletcher-Watson et al. (2008) found that observers mostly fixated on faces during their initial stages of viewing (first to third fixations). Later, their fixations on objects looked at by people in the scenes increased (fourth fixation and later). Such second-stage fixation directed by a gazing face (i.e., attentional guidance by gaze direction and attentional priority to hands) must be crucial for efficient perception of others' behavior and intentions.

Given these findings, the interactions among objects such as faces, hands, and objects manipulated by hands prove critical to understanding the attentional mechanism supporting human scene perception. Attentional priorities among objects in natural scenes are not absolute but relative: Visual orienting may be well understood as biased competition among objects (Desimone, 1998). The facial advantage may be absent without other competing objects (Ro, Russell, & Lavie, 2001). In the absence of faces, bodies, salient objects, and text strings in scenes may draw attention as effectively as faces (Bindemann et al., 2010; Cerf, Frady, & Koch, 2009). Attention to hands is better understood in relation to other competitive objects. It

should be noted, however, that the present study failed to show that hands have attentional priority as strong as faces have (Fig. 2). It seems true nonetheless that the attentional priority for faces is exceptionally absolute.

In most cases, faces drew attention at the very initial stages of viewing, resulting in fast person detection. After that, the gaze direction of the depicted face or the interacting hands led the observer's attention to clues toward understanding the person's goal-directed behaviors and intentions. Combinations of multiple objects in a scene generated this pattern of gaze shift. Although hands and other body parts held a lower attentional priority than faces, *combinations* of body parts (hand and object, face and hand, etc.) are crucial for understanding how humans perceive other humans in natural scenes.

Open practices statement The data and materials for all experiments are available on request. None of the experiments was preresistered.

References

Bindemann, M., Scheepers, C., Ferguson, H. J., & Burton, A. M. (2010). Face, body, and center of gravity mediate person detection in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 1477–1485.

Birmingham, E., Bischof, W. E., & Kingstone, A. (2008a). Gaze selection in complex social scenes. *Visual Cognition*, *16*, 341–355.

Birmingham, E., Bischof, W. E., & Kingstone, A. (2008b). Social attention and real-world scenes: The roles of action, competition and social content. *Quarterly Journal of Experimental Psychology*, *61*, 986–998.

Bracci, S., Ietswaart, M., Peelen, M. V., & Cavina-Pratesi, C. (2010). Dissociable neural responses to hands and non-hand body parts in human left extrastriate visual cortex. *Journal of Neurophysiology*, *103*, 3389–3397.

Bracci, S., & Peelen, M. V. (2013). Body and object effectors: The organization of object representations in high-level visual cortex reflects body–object interactions. *Journal of Neuroscience*, *33*, 18247–18258.

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video. *Psychological Science*, *10*, 243–248.

Cerf, M., Frady, E. P., & Koch, C. (2009). Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of Vision*, *9*(12), 10:1–15.

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society B*, *353*, 1245–1255. https://doi.org/10.1098/rstb.1998.0280

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience*, *4*, 2051–2062.

Downing, P. E., Bray, D., Rogers, J., & Childs, C. (2004). Bodies capture attention when nothing is expected. *Cognition*, *93*, B27–B38.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*, 2470–2473.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition*, *6*, 509–540.

Dukewich, K. R., Klein, R. M., & Christie, J. (2008). The effect of gaze on gaze direction while looking at art. *Psychonomic Bulletin & Review*, *15*, 1141–1147. https://doi.org/10.3758/PBR.15.6.1141

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception*, *37*, 571–583.

Fletcher-Watson, S., Leekam, S. R., Benson, V., Frank, M. C., & Findlay, J. M. (2009). Eye-movements reveal attention to social information in autism spectral disorder. *Neuropsychologia*, *47*, 248–257.

Green, C., & Hummel, J. E. (2006). Familiar interacting object pairs are perceptually grouped. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1107–1119. https://doi.org/10.1037/0096-1523.32.5.1107

Harel, J. (2012). A saliency implementation in MATLAB. Retrieved from http://www.klab.caltech.edu/~harel/share/gbvs.php

Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Research*, *45*, 1707–1724. https://doi.org/10.1016/j.visres.2004.12.021

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*, 1254–1259.

Johnson-Frey, S. H., Maloof, F. R., Newman-Norlund, R., Farrer, C., Inati, S., & Grafton, S. T. (2003). Actions or hand-object interactions? Human inferior frontal cortex and action observation. *Neuron*, *39*, 1053–1058.

Kano, F., & Tomonaga, M. (2009). How chimpanzees look at pictures: A comparative eye-tracking study. *Proceedings of the Royal Society B*, *276*, 1949–1955.

Kim, J. G., Biederman, I., & Juan, C.-H. (2011). The benefit of object interactions arises in the lateral occipital cortex independent of attentional modulation from the ipsilateral sulcus: A transcranial magnetic stimulation study. *Journal of Neuroscience*, *31*, 8320–8321. https://doi.org/10.1523/JNEUROSCI.6450-10.2011

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(ECVP Abstract Suppl), 14.

Langton, S. R. H., Law, A. S., Burton, A. M., & Schweinberger, S. R. (2008). Attention capture by faces. *Cognition*, *107*, 330–342.

Mayer, K. M., Vuong, Q. C., & Thornton, I. M. (2015). Do people "pop out"? *PLoS ONE*, *10*, e0139618. https://doi.org/10.1371/journal.pone.0139618

Morrisey, M. N., & Rutherford, M. D. (2013). Do hands attract attention? *Visual Cognition*, *21*, 647–672.

Op de Beeck, H. P., Brants, M., Baeck, A., & Wagemans, J. (2010). Distributed subordinate specificity for bodies, faces, and buildings in human ventral visual cortex. *NeuroImage*, *49*, 3414–3425.

O'Toole, A. J., Phillips, P. J., Weimer, S., Roark, D. A., Ayyad, J., Barwick, R., & Dunlop, J. (2011). Recognizing people from dynamic and static faces and bodies: Dissecting identity with a fusion approach. *Vision Research*, *51*, 74–83.

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience*, *8*, 636–648.

Perini, F., Caramazza, A., & Peelen, M. V. (2014). Left occipitotemporal cortex contributes to the discrimination of tool-associated hand actions: fMRI and TMS evidence. *Frontiers in Human Neuroscience*, *8*, 591:1–10. https://doi.org/10.3389/fnhum.2014.00591

Pinsk, M., Desimone, K., Moore, T., Gross, C., & Kastner, S. (2005). Representations of faces and body parts in macaque temporal cortex: A functional MRI study. *Proceedings of the National Academy of Sciences*, *102*, 6996–7001.

Rice, A., Phillips, P. J., Natu, V., An, X., & O'Toole, A. J. (2013). Unaware person recognition from the body when face identification fails. *Psychological Science*, *24*, 2235–2243.

Ro, T., Friggel, A., & Lavie, N. (2007). Attentional biases for faces and body parts. *Visual Cognition*, *15*, 322–348.

Ro, T., Russell, C., & Lavie, N. (2001). Changing faces: A detection advantage in the flicker paradigm. *Psychological Science*, *12*, 94–99.

Robbins, R., & Coltheart, M. (2012). The effects of inversion and familiarity on face versus body cues to person recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 1098–1104.

Simhi, N., & Yovel, G. (2016). The contribution of the body and motion to whole person recognition. *Vision Research*, *122*, 12–20.

Sogo, H. (2013). GazeParser: An open-source and multiplatform library for low-cost eye tracking and analysis. *Behavior Research Methods*, *45*, 684–695.

VanRullen, R. (2006). On second glance: still no high-level pop-out effect for faces. *Vision Research*, *46*, 3017–3027.

Vuilleumier, P., George, N., Lister, V., Armony, J., & Driver, J. (2005). Effects of perceived mutual gaze and gender on face processing and recognition memory. *Visual Cognition*, *12*, 85–101.

Wang, Y., Newport, R., & Hamilton, A. F. (2011). Eye contact enhances mimicry of intransitive hand movement. *Biology Letters*, *7*, 7–10.

Zwicker, J., & Võ, M. L.-H. (2010). How the presence of persons biases eye movements. *Psychonomic Bulletin & Review*, *17*, 257–262. https://doi.org/10.3758/PBR.17.2.257