Action recognition in the visual periphery

Laura Fademrecht　　　Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Isabelle Bülthoff　　　Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Stephan de la Rosa　　　Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Recognizing whether the gestures of somebody mean a greeting or a threat is crucial for social interactions. In real life, action recognition occurs over the entire visual field. In contrast, much of the previous research on action recognition has primarily focused on central vision. Here our goal is to examine what can be perceived about an action outside of foveal vision. Specifically, we probed the valence as well as first level and second level recognition of social actions (handshake, hugging, waving, punching, slapping, and kicking) at 0° (fovea/fixation), 15°, 30°, 45°, and 60° of eccentricity with dynamic (Experiment 1) and dynamic and static (Experiment 2) actions. To assess peripheral vision under conditions of good ecological validity, these actions were carried out by a life-size human stick figure on a large screen. In both experiments, recognition performance was surprisingly high (more than 66% correct) up to 30° of eccentricity for all recognition tasks and followed a nonlinear decline with increasing eccentricities.

Introduction

Recognition of human actions is crucial for social interaction. So far, most studies have investigated the visual mechanisms underlying action recognition at fixation (central vision) and largely ignored peripheral vision. However, in real life, we are aware of actions happening not only in central vision but also in the visual periphery. For example, in a conversational setting, we are still aware of our partner's hand movements despite focusing on his face. The purpose of the present study is to examine the recognition of social actions throughout the visual field, that is, in the central and peripheral regions of the retina.

Many of the studies investigating visual recognition of bodily movements within the central vision field have shown that humans are able to read a large range of information from biological motion (for comprehensive reviews, see Blake & Shiffrar, 2007; Giese, 2013), for example, the actor's identity (Cutting & Kozlowski, 1977; Loula, Prasad, Harber, & Shiffrar, 2005), intention (Runeson & Frykholm, 1983), or sex (Barclay, Cutting, & Kozlowski, 1978; Kozlowski & Cutting, 1977). Yet, everyday social interactions also require humans to be exquisite at recognizing actions. For example, the generation of an appropriate complementary action requires the observer to determine whether the interaction partner is carrying out a punch or a handshake. Only a few studies have investigated the recognition of social actions. They have shown that their recognition is sensitive to the temporal synchrony and the semantic relationship of the interaction partner's actions (Manera, Becchio, Schouten, Bara, & Verfaillie, 2011; Neri, Luu, & Levi, 2006). In addition, social action recognition is also sensitive to viewpoint (de la Rosa, Mieskes, Bülthoff, & Curio, 2013) and to the social context in which an action is embedded (de la Rosa, Streuber, Giese, Bülthoff, & Curio, 2014; Streuber, Knoblich, Sebanz, Bülthoff, & de la Rosa, 2011). Moreover, we can recognize the same action on several cognitive abstraction levels (first level: e.g., handshake; second level: e.g., greeting; de la Rosa et al., 2015).

Action recognition in the visual periphery has received little attention. The few existing studies, all of which used point light stimuli, have mainly focused on the detection and the direction discrimination of locomotive actions (e.g., walking, running) at eccentricities up to 12° (near periphery). Their results show that these actions can be readily detected at these

eccentricities, although there was always a disadvantage in the periphery compared with central vision (Ikeda, Blake, & Watanabe, 2005; Ikeda & Cavanagh, 2013; B. Thompson, Hansen, Hess, & Troje, 2007).

There are several reasons to assume that the role of peripheral vision with regard to action recognition goes beyond the detection of biological motion and the discrimination of the direction of an action. Previous research suggests that at least two other important aspects of an action could be detected in the periphery, namely, we can judge its emotional valence and classify it at various abstraction levels. As for valence, face recognition research suggests that affective face information can be readily recognized in the visual periphery (Bayle, Henaff, & Krolak-Salmon, 2009; Bayle, Schoendorff, Hénaff, & Krolak-Salmon, 2011; Rigoulot, D'Hondt, Honoré, & Sequeira, 2012). With regard to actions, the recognition of their emotional valence in the visual periphery would, for example, allow an early detection of a threatening action. In terms of cognitive abstraction levels, previous research has shown that action categorization occurs on several abstraction levels (de la Rosa et al., 2015). For instance, participants could describe a handshake action as a greeting or on a more detailed level as a handshake. The former is referred to as *recognition at the second level* and the latter as *recognition at the first level*. These different recognition levels result in different levels of recognition performance. For the description of these recognition levels, we refer to the specificity of the actions. That means that for the more specific group of actions (e.g., handshake), we use the term *first level*, and for the more general group of actions (e.g., greeting) we will use the term *second level*. In congruence with the object recognition literature, in which the second level is described as basic level and the first level as subordinate level (Rosch, Mervis, Gray, Johnson, & Boyes-braem, 1976), actions are recognized more accurately and faster at the second than at the first level (de la Rosa et al., 2015).

In two experiments, we examined the visual recognition of social actions in the periphery. In Experiment 1, we examined the recognition of dynamic actions with respect to their valence, first and second level of recognition. In Experiment 2, we investigated valence, first level and second level recognition for static and dynamic actions.

Experiment 1

Our aim was to examine valence (positive vs. negative), first level (e.g., handshake), and second level

(e.g., greeting) action recognition over a large portion of the visual field (fixation, near periphery, and far periphery). To mimic realistic viewing scenarios, we used dynamic actions (i.e., movies) and kept the size of the life-size actor constant across the visual field instead of adjusting the stimulus size to compensate for the reduced resolution in the periphery (cortical magnification; Cowey & Rolls, 1974; Daniel & Whitteridge, 1961; Rovamo & Virsu, 1979).

Methods

Participants

We recruited 45 participants (18 men, 27 women) from the local community of Tübingen. All participants received monetary compensation for their participation. Their age ranged from 19 to 53 years ($M = 27.1$). The participants had normal vision or corrected their visual acuity using contact lenses. Participants gave their written, informed consent form prior to the experiment. The study was conducted in line with Max Planck Society policy and had been approved by the University of Tübingen ethics committee.

Apparatus

Stimuli were presented on a large panoramic screen with a half-cylindrical virtual reality projection system (Figure 1). The wide screen display amounts to 7 m in diameter and 3.2 m in height (230° horizontally, 125° vertically). Six LED DLP projectors ($1,920 \times 1,200$, 60 Hz; EYEVIS, Reutlingen, Germany) were used to display the stimuli against a gray background on the screen. The geometry of the screen can be described as a quarter-sphere. The visual distortions of the display caused by the curved projection screen were compensated with the use of warping technology software (NVIDIA, Santa Clara, CA). At all eccentricities the screen was 2 m away from the subject, and the stimuli were presented at a virtual distance of 3 m. Participants sat on a stool in front of a desk in the middle of the quarter-spherical arena. They placed their head on a chin and forehead rest, which was mounted on the desk (see Figure 1). During each experimental trial, they were required to keep their eyes focused on a gray fixation cross placed on the screen straight ahead of them. This position of the cross was defined as 0° position. An eye tracker (Eyelink II, SR Research Ltd., Ottawa, Ontario, Canada) was used to control for their eye movements. When the stick figure was presented at 0°, it was presented behind the cross. The Unity 3D (Unity Technologies, San Francisco, CA) game engine in combination with a custom written control script was used to control the presentation of the stimuli and to collect responses.
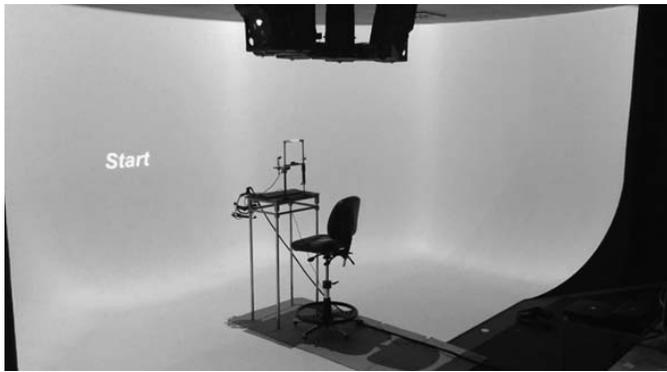
Figure 1. Projection system: large panoramic screen with chair, chin rest, and eye tracker.

Stimuli

Actions were recorded via motion capture using Moven Suits (XSens, Enschede, the Netherlands). The Xsens MVN Suits consist of 17 inertial and magnetic sensor modules that are placed in an elastic Lycra suit worn by the actor. The sampling rate was 120 Hz. Three actions with positive emotional valence (handshake, hugging, and waving) and three actions with negative valence (slapping, punching, and kicking) were acted out each by six different lay actors (three male, three female). Every action was repeated six times by each actor, leading to 216 stimuli in total. The actions lasted between 800 and 1500 ms, and each action started with the actor standing in a neutral position (i.e., with arms aligned with the body) and ended with the peak frame of the action. The peak frame of an action was determined as the point in time just before the actor started moving back to the neutral position.

The motion data were mapped onto a gray life-size stick figure avatar (avatar height: 170 cm, about 32° visual angle). The position of the stick figure avatar in the visual field (on the screen) was determined by the position midway between both hips. The stimuli were always oriented toward the participant at any position on the screen and were always presented along the same latitude (i.e., on the same horizontal axis). A stick figure (see Figure 1) was used instead of a full-fleshed avatar in order to prevent any other visual cues such as appearance or gender from influencing the participant's decisions. Furthermore, using a stick figure had the advantage that we did not have to record facial movement information (e.g., expression and gaze) and hand and feet motions. We favored the use of a stick figure over a dynamic point-light display because the sparse structure of the latter might unduly hinder recognition because of the decreasing spatial resolution of the visual system toward the periphery.

Procedure and design

At the beginning of the experiment, participants were informed about the following experimental procedure (Figure 2). Each trial began with the presentation of a fixation cross, and the eye tracker started to record the eye movements. Participants were told to fixate the cross (trials with a gaze shift larger than 2° were discarded from all analyses). The stick figure appeared at one of nine positions (−60°, −45°, −30°, −15°, 0°, 15°, 30°, 45°, or 60°) in the participant's visual field. Participants were instructed to answer one of the following three questions in a between-subject design. They answered either the question, "What action did you see?" in order to identify the action they had seen (first level task), or they answered the question, "Was the action a greeting or an attack?" meaning that they categorized the actions at the second level (second level task), or they answered the question, "Was the action positive or negative?" to evaluate the emotional valence of the viewed action (valence task). There were two answer options in the valence task (i.e., positive or negative) and in the second level task (i.e., greeting or attack). There were six answer options in the first level task (handshake, hugging, waving, kicking, punching, or slapping). Participants were asked to answer as quickly and as accurately as possible. The answer could be given as soon as the stick figure appeared on the screen. When participants did not respond before the end of the animation sequence, a prompt appeared on the screen, displaying the question and the predefined response keys on a keyboard (1 or 0 on the keyboard for the second level and the valence task; 1, 2, 3, 8, 9, 0 on the keyboard for the first level task). Three of the actions had a positive emotional valence (handshake, hugging, waving), and three had a negative emotional valence (kicking, punching, slapping). Each of these six actions was presented 100 times. We manipulated the eccentricity of the stick figure, so that it appeared at 0°, 15°, 30°, 45°, or 60° away from the fixation cross (Figure 1). The stick figure appeared randomly either on the left or on the right side of the screen for positions other than 0°. The actions (and hence the valence) and their positions on the screen were counterbalanced within each task with each action presented 20 times at each location. This resulted in a total of 600 trials per task (20 repetitions × five position × six actions). In 600 trials, the 216 stimuli were shown 2.8 times on average. Actions and positions were in a different random order for every participant.

Each recognition task was performed by a separate group of 15 participants. Hence, recognition task was a between-subjects factor, and position, action, and valence were within-subjects factors. At the beginning of an experiment, participants received a short training to get used to the setup and the task. In the second level and valence tasks, this training lasted for 10 trials; in
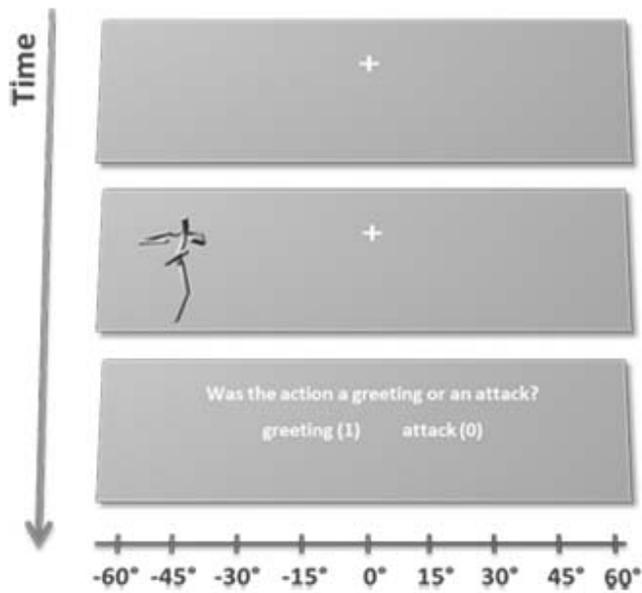
Figure 2. Sequence of events within a trial.

the first level task, participants received a longer training phase of 20 trials to learn the response key–action associations. The stimuli used in the training trials were different from the stimuli in the test trials.

Results

Accuracy and reaction times served as measures for recognition performance, and their results are presented here separately. In less than 0.8% of the trials, participants failed to fixate the cross in the middle of the screen. These trials were discarded. For the analysis, we collapsed the data of the right and left visual field. Reaction time data were filtered for outliers, and reaction times less than 200 ms and greater than 4000 ms were discarded (0.6% of the trials).

Reaction times

Only reaction times for correct responses were considered in this analysis. The mean reaction time overall was 1181 ms ($SE = 16$ ms). Participants' reaction times increased with eccentricity for each recognition task and were task dependent (Figure 3). We ran a mixed-effects model with recognition task (first level, second level, valence) and eccentricity (0°, 15°, 30°, 45°, 60°) as fixed factors and participant as random factor. The slope for eccentricity was fitted in a by-participant fashion. The results showed a significant main effect of eccentricity, $F(1, 177) = 184.51$, $p < 0.0001$, and a significant main effect of

recognition task, $F(2, 42) = 16.56$, $p < 0.001$. The results suggest that reaction times were dependent on the stimulus position in the visual field and on the recognition task. We examined the effect of task with pairwise t tests using a Bonferroni correction for multiple comparisons. These showed that participants answered faster in the valence recognition task than in the second level task (valence vs. second level: $t_{paired} = 5.42$, $df = 147.57$, $p < 0.001$). They showed the longest reaction times in the first level task (second level vs. first level: $t_{paired} = 7.12$, $df = 130.97$, $p < 0.001$). This worse performance in the first level task might be due to the larger number of response options (six response options) compared with the valence and the second level task (two response options). The two-way interaction was significant, $F(2, 177) = 3.51$, $p = 0.032$, which shows that for each recognition task, reaction times increased differently with eccentricity. The significant interaction between recognition task and eccentricity was examined using Dunnett's test. We were interested in the position at which recognition performance in the periphery started to differ significantly from foveal vision. We therefore compared all peripheral positions to 0° eccentricity. For the second level and the valence task, reaction times at fixation and in the periphery did not differ significantly from each other up to and including 45° eccentricity (all p values higher than 0.1). Thus, the only significant difference was between 0° and 60° (second level: $t_{paired} = 4.07$, $p < 0.001$; valence: $t_{paired} = 3.29$, $p < 0.01$), indicating that there was no significant increase of reaction times before testing at 60° eccentricity. For the first level task, there was even no significant difference to the reaction times at fixation for all tested eccentricities (all p values higher than 0.1).
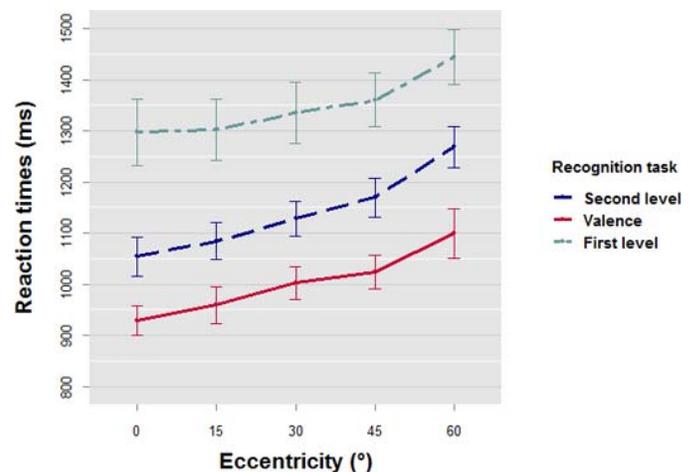

Figure 3. Means and standard errors of reaction times for the three recognition tasks as a function of eccentricity.
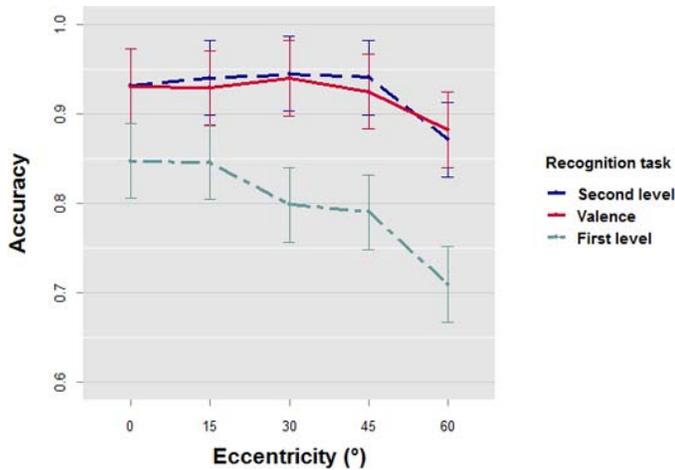
Figure 4. Means and 95% confidence intervals of response accuracy for the three recognition tasks as a function of eccentricity.

Accuracy

To account for the fact that the first level task had six response options whereas the second level and the valence tasks had only two, we corrected the accuracy results statistically for guessing according to Macmillan and Creelman (2005, p. 252):

$$c = [m \cdot p(c) - 1]/(m - 1) \cdot 100$$

This formula gives the accuracy corrected for guessing in percent c. The parameter $p(c)$ is the probability of a correct response, m is the number of response options in a given task (for the valence and the second level task $m = 2$; for the first level task, $m = 6$).

The mean accuracy was well above chance level for each task over all tested eccentricities (overall accuracy: $M = 0.88$, $SE = 0.005$). Figure 4 shows that recognition performance decreased with eccentricity in all tasks and that accuracy was lower for the first level task than for the second level and the valence task. Furthermore, the decline of performance with eccentricity seems stronger for the first level task than for the two other tasks. A mixed-effects model with recognition task (first level, second level, valence) and eccentricity (0°, 15°, 30°, 45°, 60°) as fixed factors and a random slope for eccentricity that was fitted in a by-participant fashion revealed a significant main effect of eccentricity, $F(1, 177) = 57.32$, $p < 0.001$, indicating that recognition performance decreased with increasing eccentricity of the stimulus position in the visual field. There was also a significant main effect of recognition task, $F(2, 42) = 29.47$, $p < 0.0001$, showing that participants' accuracy depended on the task requirements. The significant two-way interaction of task and eccentricity, $F(2, 177) = 9.26$, $p < 0.0001$, suggests that eccentricity affected recognition in the three tasks differently. In Figure 4, we plotted the 95% confidence intervals. This illustrates the significant

differences between the three tasks at the different positions. For the 0° and 15° positions, there is no performance difference between the tasks, whereas from 30° onward, the first level task always leads to lower accuracy rates than the second level and the valence tasks, whose data did not differ from each other at any position. Between the valence and the second level task, we found no difference in accuracy. We used a Dunnett's test for each recognition task to compare the recognition performance at the peripheral positions with the performance at fixation. In all three recognition tasks, the recognition performance up to 45° eccentricity did not differ significantly from the performance at fixation, thus indicating that the decline of recognition performance starts after 45° (Dunnett's test was significant only for comparisons between 0° and 60° in all recognition tasks: valence, $t = -2.57$, $p = 0.04$; second level, $t = -3.5$, $p < 0.01$; first level, $t = -5.12$, $p < 0.001$). Figure 4 in combination with the statistical analysis indicates a nonlinear relationship of recognition performance with eccentricity for all three recognition tasks.

Discussion

In this experiment, we tested human action recognition from central vision up to 60° eccentricity in three different recognition tasks (first level, second level, and the emotional valence). Reaction times in all three tasks increased with eccentricity. Moreover, participants were fastest in the valence task and slowest in the first level task, thus confirming previous findings (de la Rosa et al., 2015). The accuracy data also indicate that all tasks get harder with increasing eccentricity. The better recognition performance observed in foveal vision for second level compared with first level categorization therefore seems to extend to peripheral vision. In addition, participants showed the best performance in the valence task. For this task, accuracy is as high as in the second level task, whereas reaction times are shorter than in both other tasks. The significant interaction between task and eccentricity seems to be owed to the steeper decline in recognition performance in the first level task compared with second level and valence tasks. A simple explanation for this pattern is that more detailed visual information is needed for the recognition at the first level. This type of information might not be accessible because of the sparse resolution in the visual periphery.

Remarkably, accuracy declined nonlinearly with increasing eccentricity but remained above chance level for all tested eccentricities. The former was partly unexpected because previous research examining recognition of static objects in the periphery (Jebara, Pins, Despretz, & Boucart, 2009; Thorpe, Gegenfurt-

ner, Fabre-Thorpe, & Bülthoff, 2001) reported a linear decline of recognition performance with eccentricity. Can the motion energy in our dynamic stimuli account for this difference? The literature about motion perception in peripheral vision describes a rather linear decline with eccentricity for first- and second-order motion. Few studies describe a nonlinear relationship (Tynan & Sekuler, 1982). In this study, the authors measured motion detection thresholds up to 30° eccentricity and found a nonlinear increase of detection thresholds. However, it is important to note that the motion patterns induced by limb movement in our stimuli are more complex (e.g., they consist of many more movement orientations in three-dimensional [3D] space) than the ones employed in previous studies with low-level motion stimuli (first- and second-order motion). Hence, we cannot rule out that participants might have relied on the additional motion cues in our stimuli to maintain a high recognition performance far into the periphery. If motion cues were responsible for the nonlinear decline in our study, then presenting static action images instead of action movies should result in a linear decline of performance with eccentricity. To test this hypothesis, we conducted a second experiment in which we compared the recognition of action movies with the recognition of static representations of actions (images).

Experiment 2

Experiment 2 investigated the influence of motion information on the recognition of social actions in the visual periphery. We changed the experimental methods to overcome two important shortcomings of Experiment 1. First, the three recognition tasks in Experiment 1 had different numbers of response options (six response options in the first level task and two response options in the valence and second level tasks). The larger number of response options in the first level task might have been responsible for the slower reaction times and lower accuracy in that task. To avoid this problem, we changed the experimental design in such a way that all recognition tasks had two response options. Specifically, participants were presented with one action at a time and had to indicate whether the presented action matched a predefined first level (e.g., punching), second level (e.g., greeting), or valence (e.g., positive). Hence, all three tasks relied on a yes-no task. Yes-no tasks have been frequently used for the investigation of visual object categorization, and it has been shown that switching from an *n*-alternative forced-choice (n-AFC) task (with $n > 2$) to a yes-no task does not change the overall pattern of

the results in object categorization tasks (de la Rosa, Choudhery, & Chatziastros, 2011; Grill-Spector & Kanwisher, 2005). Hence, in Experiment 2, we switched to a yes-no task for obtaining a more fair comparison of the different recognition tasks. Second, positive and negative valence actions were associated with different motion energies in the stimuli of Experiment 1. Therefore, participants might have defaulted to a recognition strategy that relied on simply assessing the amount of motion energy in the second level and valence recognition task. Therefore, we added distractor actions in Experiment 2 that had a motion energy similar to the actions described in Experiment 1 but did not show any meaningful actions. To create these actions, we remapped the arm motion onto the legs and vice versa.

Methods

We used the same methods as in Experiment 1 except for the following.

Participants

We recruited 19 participants from the local community of Tübingen (nine men, ten women). The ages ranged from 20 to 39 years (mean = 26.1).

Stimuli

In addition to the action stimuli of Experiment 1, we created distractor stimuli in the following two ways: Either we remapped the left and right arm movements onto the left and right legs and vice versa or we mapped the left leg movement onto the right arm and vice versa. We will refer to these distractor stimuli as remapped distractor stimuli. Importantly, no action could be recognized from these actions, thereby rendering them meaningless.

Procedure and Design

We measured the recognition of dynamic and static actions in two separate experimental sessions (testing order was counterbalanced across participants). Each experimental session consisted of 10 experimental conditions (two for the valence task, two for the second level task, and six for the first level task). At the beginning of each experimental condition, participants received verbal instructions about the question they had to answer in that condition. Each question probed the recognition of a different action target. The questions, "Was the action positive?" or "Was the action negative?" probed valence recognition, and the questions, "Was the action a greeting?" and "Was the

action an attack?" measured second level recognition. We used the following six questions to measure first level recognition: "Was the action a handshake?" "Was the action a hug?" "Was the action a wave?" "Was the action a kick?" "Was the action a punch?" and "Was the action a slap?" Participants always had the response options "yes" (for the target action) and "no" (for nontargets) for each question in the static and the dynamic experimental sessions. For remapped distractors, a correct response was "no" to all questions. An experimental trial started with the presentation of the fixation cross at 0°, and the stick figure avatar appeared at one of the nine positions in the participant's visual field as described in the first experiment. The answer could be given as soon as the stick figure appeared on the screen. When participants did not respond before the end of the animation sequence, a prompt appeared on the screen, displaying the question and the predefined response keys on a keyboard. In each experimental condition, 50% of the trials showed the target action, and the remaining trials showed distractors. Fifty percent of the distractors were remapped distractor stimuli, and the remaining were nontarget actions that were not remapped. For example, if the target was "positive actions," 50% of the experimental trials showed the three positive actions as target (hugging, waving, handshake), 25% showed remapped distractors derived from the positive actions, and 25% of the trials showed the three actions with negative valence (kicking, punching, slapping). The testing order of the 10 experimental conditions was randomized across participants. In each condition, each target action was presented 12 times at each location of each hemi-field. The valence and second level tasks had 480 trials (12 repetitions × two target presence [present vs. absent] × five locations × two hemi-fields [left or right side of the visual field] × two questions [for example, "Was the action positive?"/"Was the action negative?"]). The first level task had 80 trials for each of the six questions (for example, "Was the action a handshake?" and so on; four repetitions × two target presence [present vs. absent] × five locations × two hemi-fields). The two experimental sessions probing the recognition of static and dynamic actions were carried out on different days. This resulted in a total of 2,880 trials per participant ([two × 480 trials in the second level and valence condition + 480 trials in the first level condition] × two sessions). Recognition task, position, and motion type (static vs. dynamic) were within-subjects factors.

Results

We calculated the sensitivity (d′) according to Macmillan and Creelman (2005) as a measure of

recognition performance. One percent of the trials were excluded because of deviation from fixation. Reaction times for correct target identification and sensitivity are evaluated separately. Reaction time data were filtered for outliers, and reaction times below 200 ms and above 3500 ms were discarded (0.2% of the trials).

Reaction times

Participants' reaction times increased with eccentricity for each task and both motion types (Figure 5). The mean reaction time for the whole experiment was 894 ms ($SE = 2$ ms) calculated over all data in all tasks. We analyzed only the reaction times for correct target identification. We used a mixed-effects model with recognition task (first level, second level, valence) and eccentricity (0°, 15°, 30°, 45°, 60°) and motion type (static, dynamic) as fixed factors and a random slope for eccentricity that was fitted in a by-participant fashion to investigate the reaction times. To examine the relationship between the reaction time performance and eccentricity, we treated eccentricity as a continuous variable. We found a significant main effect of recognition task, $F(2, 538) = 45.65$, $p < 0.001$. In the first level task, participants answered with shorter reaction times ($M_{RT} = 843$ ms, $SE = 1$) than in the second level task ($M_{RT} = 883$ ms, $SE = 1$) and in the valence task ($M_{RT} = 937$ ms, $SE = 2$). The significant main effect of eccentricity, $F(1, 538) = 87.51$, $p < 0.001$, indicates that participants' reaction times were dependent on the stimulus position and increased with increasing eccentricity. The motion type had a significant main effect on the reaction times as well, $F(1, 538) = 6.35$, $p = 0.01$; participants showed shorter reaction times for the static condition than for the dynamic condition, although this difference did depend on the recognition task, as the significant interaction between motion type and recognition task, $F(2, 538) = 3.86$, $p = 0.02$, shows. Pairwise t tests, using a Bonferroni correction for multiple comparisons, showed that reaction times for dynamic and static stimuli differed significantly from each other only in the first level task ($t_{paired} = 4.01$, $df = 94$, $p < 0.001$). All other comparisons and interactions were nonsignificant (all p values > 0.05).

Sensitivity (d′)

A cursory look at the graph (Figure 6) indicates that, for dynamic stimuli, participants were always clearly able to discriminate between target and distractor actions at all probed locations, as indicated by d′ values higher than 0, whereas for static stimuli, this was true only up to 30° eccentricity. A mixed-effects model with recognition task (first level, second level, valence) and eccentricity (0°, 15°, 30°, 45°, 60°)
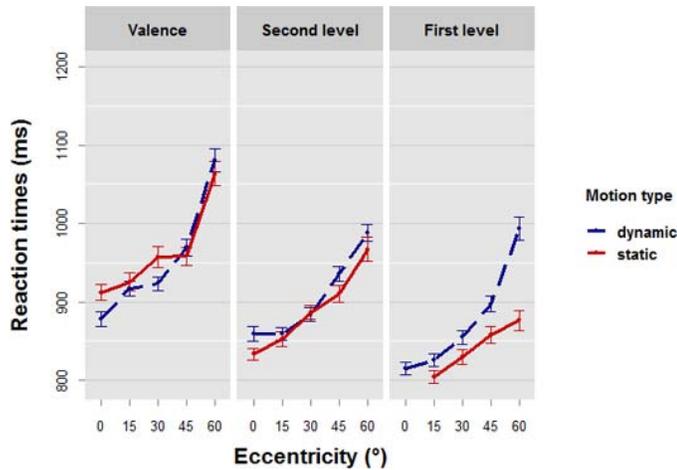
Figure 5. Means and standard errors of the reaction times as a function of eccentricity for static and dynamic action stimuli in the three recognition tasks. (Note that the scales of the *y* axis differ in Experiments 1 and 2.)
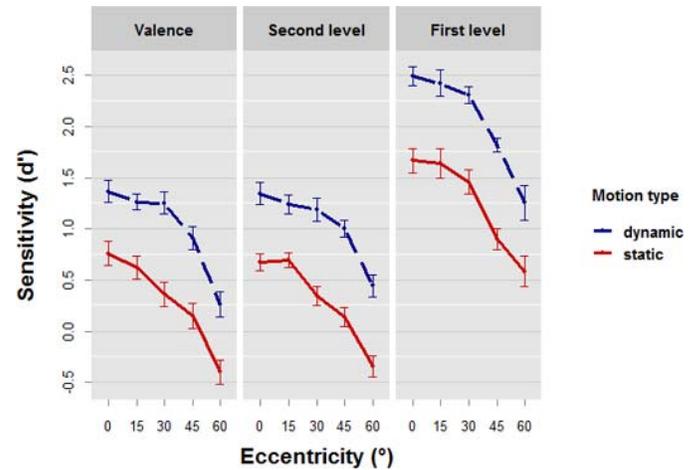

Figure 6. Means and standard errors of sensitivity (d′) as a function for eccentricity for static and dynamic action stimuli in the three recognition tasks.

and motion type (static, dynamic) as fixed factors and a random slope for eccentricity that was fitted in a by-participant fashion shows a significant main effect of recognition task, $F(2, 540) = 350.46$, $p < 0.0001$. In the first level task ($M_{d'} = 1.66$; $SE = 0.06$), participants reached significantly higher d′ values than in the second level ($M_{d'} = 0.67$; $SE = 0.05$) and in the valence task ($M_{d'} = 0.66$; $SE = 0.05$; t test: valence vs. first level task, $t_{paired} = 12.83$, $df = 374.84$, $p < 0.001$; second level vs. first level task, $t_{paired} = 13.16$, $df = 365.22$, $p < 0.001$; second level vs. valence task, $t_{paired} = -0.24$, $df = 374.48$, $p < 0.811$). This finding indicates a better recognition performance in the first level task than in the two other recognition tasks. The main effect of eccentricity was significant as well, $F(1, 540) = 100.55$, $p < 0.001$. The mean d′ averaged over all three tasks and the two conditions is decreasing with eccentricity, starting with a mean d′ of 1.38 ($SE = 0.07$) at fixation and ending with a mean d′ of 0.3 ($SE = 0.07$) at 60°. We examined the main effect of eccentricity using Dunnett's test. Sensitivity values at all peripheral positions were compared with that at fixation. We found a significant difference between 0° and 45° ($t_{paired} = -5.55$, $p < 0.001$) and as well for 0° and 60° ($t_{paired} = -10.64$, $p < 0.001$). These results indicate that the decline of recognition performance starts after 30° eccentricity. The significant main effect of motion type, $F(1, 540) = 456.86$, $p < 0.001$, shows that response accuracy is also sensitive to the experimental condition (static or dynamic), resulting in a mean d′ of 0.62 ($SE = 0.05$) for the static condition and a mean d′ of 1.37 ($SE = 0.05$) in the dynamic condition. Thus, dynamic target stimuli are better discriminated from distractors than static target stimuli. All higher-order interactions were nonsignificant (all *p* values > 0.05), including the two-

way interaction between motion type and eccentricity, $F(1, 540) = 0.58$, $p = 0.45$.

Assessing the change of recognition performance with eccentricity

To assess whether the performance changed linearly with eccentricity, we examined the relationship between recognition performance and eccentricity more formally. Specifically, we fitted a power law function to the performance data of Experiment 2 for each participant, separately for the dependent variable (RT and d′) and motion type (dynamic and static). The reasons for using a power law function were twofold. First, power laws have been shown to well describe relationships between physical properties and their perception (e.g., Steven's power law). Second, these functions give also the opportunity of a linear fit (exponent would then be 1); therefore, we could directly test whether the performance declines in a linear or a nonlinear fashion with eccentricity. The fits were carried out by means of the "gfit" function in MATLAB. We fitted the following power law function:

$$f(x) = w \cdot a \cdot x^b + c$$

The parameter w indicates whether the change in performance was increasing (w = 1) or decreasing (w = −1) with eccentricity. We set w = −1 for fitting the d′ data to describe the decrease of d′ with eccentricity. Likewise, we set w = 1 for the fitting of the reaction time data to describe the increase in RT with larger eccentricities. The parameter a is the slope of the function and scales the function along the *y* axis. b is the exponent and defines the type of relationship: For a linear relationship, we expect b = 1, and for nonlinear

Motion type	Exponent b	Slope a	Intercept c	R^2
Static	2.90	89.68	841.94	0.76
Dynamic	2.75	119.47	861.95	0.92

Table 1. Mean parameters for the power law function fitted to each participant's individual reaction time data.

Motion type	Exponent b	Slope a	Intercept c	R^2
Static	3.02	1.40 e-05	0.95	0.83
Dynamic	3.09	1.29 e-05	1.70	0.81

Table 2. Mean parameters for the power law fitted on each participant's individual d′ data.

relationships, $b \neq 1$. c is the intercept of the curve and is a measure of recognition performance at fixation. Parameters a, b, and c were free to vary.

We present the results from this analysis for reaction times and d′ separately.

Reaction times

On average, the power law functions fit the data well both in the static and dynamic condition (mean R^2 = 0.76 and 0.92, respectively). To assess the linearity of the performance decrease with eccentricity, we tested the exponents against 1. The mean exponents in the dynamic (M_{exp} = 2.75, SE = 0.35) and the static condition (M_{exp} = 2.90, SE = 0.52) were significantly different from 1 (dynamic: t_{paired} = 4.94, df = 18, $p <$ 0.001; static: t_{paired} = 3.62, df = 18, p = 0.002), suggesting a nonlinear relationship between reaction time and eccentricity. There was no significant difference between the exponents of dynamic and static conditions (t_{paired} = 0.35, df = 18, p = 0.73). The mean values for the intercept c and the slope a are listed in Table 1, as well as the R^2 values for the different conditions.

Sensitivity (d′)

The power law function fit the d′ data well in both conditions (R^2 in the static condition: 0.83; R^2 in the dynamic condition: 0.81). The mean exponents in both the dynamic (M_{exp} = 3.09, SE = 0.17) and static condition (M_{exp} = 3.02, SE = 0.14) were significantly different from 1 (dynamic: t_{paired} = 12.36, df = 18, $p <$ 0.001; static: t_{paired} = 13.91, df = 18, $p <$ 0.001). Therefore, this result suggests that d′ changes in a nonlinear fashion with eccentricity. There was no significant difference (t_{paired} = 0.41, df = 18, p = 0.68) between the mean exponents of the static and dynamic conditions. The mean values for the intercept c and the slope a, as well as the R^2 values, are given for the different conditions in Table 2.

Discussion

By using dynamic and static action stimuli in Experiment 2, we examined whether it was the presence of motion in the stimuli in Experiment 1 that had led to the nonlinear decline of recognition performance with

increasing eccentricity. Although participants did not differ in terms of reaction times between static and dynamic action stimuli in the valence and the second level task, participants showed a higher sensitivity for dynamic than for static actions over all positions in the visual field and all three tasks. Importantly, the absence of an interaction between motion type and eccentricity in our analysis indicates that recognition of static and dynamic actions declines in a similar fashion with eccentricity. A significant difference between the response times to static and dynamic stimuli is observed only in the second level recognition task, which might be explained by a flooring effect in the dynamic action condition. Dynamic actions were presented as videos (1 to 2 s long), whereas our static stimuli presented each action as one image extracted at the peak of the action (see the Methods section for more details). Hence, important action information was immediately visible for static actions but not for dynamic actions and allowed faster responses. The fastest response time recorded (800 ms) for videos reflects the minimum time needed to recognize an action in our dynamic stimuli, whereas response time can be much shorter with static stimuli. In line with this idea, a minimum reaction time of 800 to 900 ms has also been found in a study by de la Rosa and colleagues (2013) using dynamic movies of other actions. Hence, we think that the larger difference in response time between static and dynamic actions found in the first level task is simply owed to a flooring effect for dynamic but not static actions. Together these findings indicate that discriminating targets from their distractors was easier for dynamic than static stimuli at all positions in the visual field. In the dynamic condition, participants were clearly able to discriminate between target actions and distractors, even in the most peripheral positions. It is important to note that they could discriminate target actions from distorted meaningless actions that were based on the motion data of the target actions. This shows that participants did not use the stimulus' motion energy as a cue to discriminate, for example, positive from negative actions. We assessed more formally the relationship between recognition performance (RT and d′) and eccentricity for dynamic and static stimuli using a power law function. This analysis indicates that both RT and d′ change in a nonlinear fashion with eccentricity. Moreover, this nonlinear decline did not vary with motion type: A nonlinear decline was

observed when using dynamic as well as static stimuli. Hence, the nonlinear decline of action recognition performance with eccentricity was unlikely because of the motion information present in the dynamic stimuli because a nonlinear decline was also observed when using static stimuli.

What other factors might explain the nonlinear decline of action recognition performance in the visual periphery? Unfortunately, most of the behavioral studies examining recognition performance in the periphery reported a linear decline and therefore provide little insights into the nonlinear nature of our results. One hypothesis that we are currently assessing in our lab is whether the perceptual field size of action-sensitive channels can account for the nonlinear decline of action recognition performance in the visual periphery. Specifically, one way to account for these results is that foveal perceptual channels extend into the periphery, thereby increasing recognition performance there. However, the reason for observing a nonlinear decline instead of a linear one is still unsettled, and what it means in terms of underlying perceptual mechanisms remains to be elucidated.

General discussion

Our results demonstrate that participants are remarkably good at recognizing actions in the periphery. Recognition performance for social actions remained above chance level up to 60° for dynamic action stimuli. Recognition of static action stimuli remained reliable up to 30° eccentricity, which indicates that up to that level of eccentricity, participants do not rely only on motion information when recognizing actions. Moreover, participants were not only able to tell the valence level of an action shown dynamically up to 60° but also its first level and second level. Hence, participants recognize much more than the emotional gist of an action even in very peripheral vision. These results parallel Thorpe and colleagues' (2001) findings, which showed that humans are very good at recognizing objects in the visual periphery (up to 70.5°). Likewise, Jebara et al. (2009) showed that participants recognized objects and faces above chance level up to 60°. Similar results have also been reported in the perception of low-level visual stimuli such as color (Naïli, Despretz, & Boucart, 2006). It has also been shown for biological motion stimuli, although there was always a disadvantage for recognition in the periphery in comparison with the visual abilities in central vision (Ikeda et al., 2005; Ikeda & Cavanagh, 2013; B. Thompson et al., 2007). Our results extend those previous findings by demonstrating that we are even able to recognize complex stimuli such as social actions in the visual periphery. Therefore, peripheral vision might play a more important role in daily social interactions than just triggering gaze saccades toward conspicuous events in the periphery. One interesting observation in both experiments was the nonlinear decline of recognition performance with eccentricity. This seems to be at variance with previous studies that report a linear decline in the recognition of static objects (Jebara et al., 2009; Thorpe et al., 2001). To examine whether motion information is at the heart of the nonlinear decline, we compared the recognition performance of dynamic and static action stimuli in Experiment 2. The results showed a nonlinear decline of recognition performance with eccentricity for both types of stimuli. Therefore, the nonlinear decline of performance with eccentricity cannot be attributed to the presence of motion information in our stimuli. Jebara et al. (2009) used smaller pictures of faces, buildings, and objects (10° visual angle) than in our study and found a linear decline, whereas Thorpe et al. (2001) also reported a linear decline, although they used very large images (39° of visual angle high, 26° across) in which all displayed animals (e.g., an insect or a tiger) had more or less the same large visual size. Our stimuli are smaller than those of Thorpe et al. (2001) and larger than those of Jebara et al. (2009). Therefore, the size of the stimuli cannot be a factor explaining why performance declines differently for objects and social actions.

Our results are also relevant for the discussion about first and second levels in action recognition (see de la Rosa et al., 2015). One key feature of second level recognition is that it is faster than first level recognition (Rosch et al., 1976). Although our results of Experiment 1 are in line with previous reports about action recognition being faster and more accurate at the second level (e.g., recognizing an action as a greeting), the results of Experiment 2 do not support this expectation. In Experiment 2, in which all tasks have been equated in terms of response possibilities, we found the shortest reaction times and the highest recognition performance for the first level task, whereas second level recognition now seemed to be the more difficult task with longer reaction times and lower recognition performance. We argue that this reversal between Experiments 1 and 2 is not due to this equalization. If the equalization of response options was responsible for this response reversal, we would expect the pattern for first and second level recognition to reverse irrespective of the stimulus type (e.g., objects or social interactions). de la Rosa and colleagues (de la Rosa et al., 2011) as well as Grill-Spector and Kanwisher (2005) have shown that equating for response options in object recognition tasks does not change the pattern of results with regard to first and second level recognition.

What might be then the reason for the reversal of first and second level recognition between Experiments 1 and 2? We suggest that this reversal might be understood in terms of current action recognition models (Bertenthal & Pinto, 1994; Giese & Poggio, 2003; Lange & Lappe, 2006; Theusner, de Lussanet, & Lappe, 2014). These theories assume that visual action information is mapped onto neuronal units that encode an action by means of many temporally ordered posture "snapshots" (action template) that encode actions akin to individual frames of a movie showing a human action. We suggest that participants might have used this template-matching mechanism more effectively in the first level task of Experiment 2 than in Experiment 1, thereby causing the reversal of the pattern. In particular, we asked participants to judge the target action in terms of a specific aspect (e.g., "Was the action a greeting: yes or no?" or "Was the action a handshake: yes or no?") in Experiment 2. Participants might have therefore benefited from top-down activation of the corresponding target action template, which resulted in matching all visual information onto this template in order to recognize the action. Such top-down influence is less efficient in the second level task than in the first level task. In the former case, visual information must be matched onto three (e.g., handshake, hugging, and waving) instead of one action template in the latter case (e.g., handshake) to recognize the action. Hence, in the assumption that participants relied on a top-down controlled template-matching strategy, one would expect first level recognition to be faster than second level recognition in Experiment 2. The same mechanisms could explain the recognition performance in Experiment 1, in which no a priori information about the target was provided. If participants relied on the same mechanism, they must have matched visual information against all six action templates in the first level recognition task. In the second level recognition task, participants could have chosen to monitor only one of the two levels (i.e., greeting or attack) and matched visual information onto the three corresponding action templates. In case there was no match, participants could have concluded that the nonchosen second level was displayed. In any case, matching visual information onto three action templates in the second level task should lead to better recognition performance than matching visual information onto six action templates in the first level recognition task. We found a decline in recognition performance that starts at smaller eccentricities in the first level task than in the second level and the valence tasks only in Experiment 1 but not in Experiment 2. This could be attributed to the fact that different action templates were needed to perform the first level task in Experiment 1. Participants needed more details to recognize the actions in order to categorize them. The

visual resolution in the periphery was not sufficient for that task. A top-down controlled template-matching mechanism could therefore, in theory, explain the reversal of first and second level recognition performance between Experiment 1 and 2. Previous literature has already shown, with behavioral and neuroimaging evidence, the strong influence top-down mechanisms (e.g., attention, goal) have on the recognition of human actions (Bülthoff, Bülthoff, & Sinha, 1998; de la Rosa et al., 2014; Grezes, 1998; Hudson et al., 2015; J. Thompson & Parasuraman, 2012).

Overall, the classification of social actions in the different recognition levels seems to be less robust than for object recognition (de la Rosa et al., 2011; Grill-Spector & Kanwisher, 2005) in the sense that changing from an n-AFC task (with $n > 2$) to a 2AFC task does not alter the overall pattern of results between the first and second level recognition tasks for object recognition. Note that participants could have defaulted to the same second level recognition strategy in the valence recognition task. The reason for this is that the actions underlying the second level recognition (attack vs. greeting) were the same actions underlying the valence levels (negative vs. positive). Despite this possibility, we find differences in performance between second level and valence recognition. This difference is suggestive of participants relying on at least partly different response strategies in these two tasks.

Furthermore, we would like to stress that second level and valence recognition in Experiments 1 and 2 were unlikely to be guided by a coarse assessment of the motion energy of the displayed action (e.g., more motion energy means negative or attack actions). When we changed the paradigm in Experiment 2 to make motion energy a much less effective cue for action classification by creating additional remapped distractor stimuli that had similar motion energy to the targets but did not show any meaningful actions, participants were still able to correctly classify the movies into their relevant categories.

To what degree could participants' performance have relied on alternative recognition such as "limb-spotting"? Because some actions are unique in the sense that they involved a unique limb (e.g., kicking action), it is possible that participants relied on monitoring the leg for the identification of the kicking action (i.e., defaulted to "limb spotting" instead of an action recognition strategy). To address the issue, we examined the sensitivity results of each action in Experiment 2 (we did not look at RT since differences in RT might be simply owed to the length of the videos). A limb-spotting strategy for kicking should lead to more accurate responses to this action in all tasks. Contrary to that expectation, kicking is an action that is recognized with an intermediate recognition performance in the second level and valence level tasks. As for

the first level task, kicking is indeed associated with the best recognition performance. However, this recognition performance is not significantly different from another action, namely, handshake ($t = 1.36$, $df = 31.17$, $p = 0.19$). Similarly, the results for hugging (the only action with bimanual movement) did not confirm the use of a spotting strategy. Therefore, we think that limb spotting contributes little to the observed effects. Moreover, if participants would solely rely on limb spotting, recognizing actions carried out by the same limb (waving, slapping, punching) should be difficult to discriminate, which should lead to reduced recognition performance. However recognition performance of these actions is well above chance level.

Using a blocked instead of a random presentation of the different conditions might account for the high recognition performance in Experiment 2 to a certain degree. Although we have no evidence on the effect of blocking versus randomizing on recognition performance, we believe that this would have little influence on the results. We think that the nature of the task (i.e., looking out for an action in an array of action movies) in Experiment 2 led participants to rely on a top-down controlled recognition strategy in which participants monitor only the action channel relevant to the task to make the judgment whether the target action had been shown. It is well conceivable that participants can quickly switch between action channels that they would like to monitor. Hence, if trials were randomized, participants were very likely to start to monitor the channel corresponding to the target action by the time they had read the target action word presented at the beginning of the trial. As a result, we would expect very little performance difference between blocked and randomized conditions. In our opinion, the largest performance change between blocked and randomized condition would be owed to a higher error rate because of the permanently switching action channels, which might lead participants to accidentally monitor the incorrect channel.

Last, we would like to point out our efforts to maintain a high ecological validity in our study for obtaining more robust results. Although one might argue that stick figures are not ideal, we think that in this initial study, we found the simplest solution to avoid distracting discrepancies between body and face. Importantly, the large curved screen combined with the correction of distortion in the display ensured that actions were displayed at equal distance from the observer in a nondistorted ecological valid fashion. The use of life-size stimuli that were not scaled with eccentricity further allowed investigation of the perception of actions in the visual periphery under more naturalistic conditions. In everyday life, the size of another person does not change across eccentricities as long as the interpersonal distance does not change. Our study lines up with recent efforts that aim at investigating action recognition under more ecologically valid viewing conditions, (e.g., Thornton, Wootton, & Pedmanson, 2014). These authors investigated the recognition of actions that were presented at various distances from the viewer and found that performance remains remarkably good even when the stimulus is moved far away along the line of sight. Here, we also find a high level of recognition despite lateral shifts of actions into the visual periphery.

Conclusion

The results of this study show that the recognition of another person's actions is well above chance level even in far periphery. In Experiment 1, participants were able to categorize dynamic actions at the first and second level and recognized their emotional valence up to 60° eccentricity. In the second experiment, we showed that the recognition performance decreased with eccentricity in a nonlinear fashion for static and dynamic actions. This indicates that the nonlinear decline is unlikely due to the motion information in the dynamic stimuli.

Keywords: peripheral vision, action perception, social interaction

Acknowledgments

References

Barclay, C. D., Cutting, J. E., & Kozlowski, L. T. (1978). Temporal and spatial factors in gait perception that influence gender recognition. *Perception & Psychophysics*, *23*, 145–152, doi:10.3758/BF03208295.

Bayle, D. J., Henaff, M.-A., & Krolak-Salmon, P. (2009). Unconsciously perceived fear in peripheral

vision alerts the limbic system: a MEG study. *PloS One*, *4*, e8207, doi:10.1371/journal.pone.0008207.

Bayle, D. J., Schoendorff, B., Hénaff, M.-A., & Krolak-Salmon, P. (2011). Emotional facial expression detection in the peripheral visual field. *PloS One*, *6*, e21584, doi:10.1371/journal.pone.0021584.

Bertenthal, B. I., & Pinto, J. (1994). Global processing of biological motions. *Psychological Science*, *5*, 221–224, doi:10.1111/j.1467-9280.1994.tb00504.x.

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, *58*, 47–73, doi:10.1146/annurev.psych.57.102904.190152.

Bülthoff, I., Bülthoff, H., & Sinha, P. (1998). Top-down influences on stereoscopic depth-perception. *Nature Neuroscience*, *1*, 254–257, doi:10.1038/699.

Cowey, A., & Rolls, E. T. (1974). Human cortical magnification factor and its relation to visual acuity. *Experimental Brain Research, 21*, 447–454, doi:10.1007/BF00237163.

Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society*, *9*, 353–356.

Daniel, P. M., & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *Journal of Physiology*, *159*, 203–221.

de la Rosa, S., Choudhery, R. N., & Chatziastros, A. (2011). Visual object detection, categorization, and identification tasks are associated with different time courses and sensitivities. *Journal of Experimental Psychology. Human Perception and Performance*, *37*, 38–47, doi:10.1037/a0020553.

de la Rosa, S., Choudhery, R. N., Curio, C., Ullman, S., Assif, L., & Bülthoff, H. H. (2014). Visual categorization of social interactions. *Visual Cognition*, *22*(9–10), 1233–1271, doi:10.1080/13506285.2014.991368.

de la Rosa, S., Mieskes, S., Bülthoff, H. H., & Curio, C. (2013). View dependencies in the visual recognition of social interactions. *Frontiers in Psychology*, *4*, 752, doi:10.3389/fpsyg.2013.00752.

de la Rosa, S., Streuber, S., Giese, M., Bülthoff, H. H., & Curio, C. (2014). Putting actions in context: Visual action adaptation aftereffects are modulated by social contexts. *PLoS One*, *9*, doi:10.1371/journal.pone.0086502.

Giese, M. A. (2013). Biological and body motion perception. In J. Wagemans (Ed.), *The Oxford handbook of perceptual organization.* (pp. 575–596). Oxford, UK: Oxford University Press.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews. Neuroscience*, *4*, 179–192, doi:10.1038/nrn1057.

Grezes, J. (1998). Top down effect of strategy on the perception of human biological motion: A PET investigation. *Cognitive Neuropsychology*, *15*, 553–582, doi:10.1080/026432998381023.

Grill-Spector, K., & Kanwisher, N. (2005). Do you know what it is as soon as you know it is there? *Psychological Science*, *16*, 152–160, doi:10.1167/8.6.512.

Hudson, M., Nicholson, T., Simpson, W. A., Ellis, R., Bach, P., Hudson, M., & Ellis, R. (2016). One step ahead: The perceived kinematics of others' actions are biased toward expected goals. *Journal of Experimental Psychology, General*, *145*, 1–7, doi:10.1037/xge0000126.

Ikeda, H., Blake, R., & Watanabe, K. (2005). Eccentric perception of biological motion is unscalably poor. *Vision Research*, *45*, 1935–1943, doi:10.1016/j.visres.2005.02.001.

Ikeda, H., & Cavanagh, P. (2013). Crowding of biological motion stimuli. *Journal of Vision*, *13*(4): 20, 1–6, doi:10.1167/13.4.20. [PubMed] [Article]

Jebara, N., Pins, D., Despretz, P., & Boucart, M. (2009). Face or building superiority in peripheral vision reversed by task requirements. *Advances in Cognitive Psychology*, *5*, 42–53, doi:10.2478/v10053-008-0065-5.

Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics*, *21*, 575–580.

Lange, J., & Lappe, M. (2006). A model of biological motion perception from configural form cues. *Journal of Neuroscience*, *26*, 2894–2906, doi:10.1523/JNEUROSCI.4915-05.2006.

Loula, F., Prasad, S., Harber, K., & Shiffrar, M. (2005). Recognizing people from their movement. *Journal of Experimental Psychology. Human Perception and Performance*, *31*, 210–220, doi:10.1037/0096-1523.31.1.210.

Manera, V., Becchio, C., Schouten, B., Bara, B. G., & Verfaillie, K. (2011). Communicative interactions improve visual detection of biological motion. *PLoS One*, *6*, e14594, doi:10.1371/journal.pone.0014594.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory. A user's guide.* Mahwah, NJ: Lawrence Erlbaum.

Naïli, F., Despretz, P., & Boucart, M. (2006). Colour recognition at large visual eccentricities in normal observers and patients with low vision. *Neuro-*

report, *17*, 1571–1574, doi:10.1097/01.wnr. 0000236864.54575.98.

Neri, P., Luu, J. Y., & Levi, D. M. (2006). Meaningful interactions can enhance visual discrimination of human agents. *Nature Neuroscience*, *9*, 1186–1192, doi:10.1038/nn1759.

Rigoulot, S., D'Hondt, F., Honoré, J., & Sequeira, H. (2012). Implicit emotional processing in peripheral vision: Behavioral and neural evidence. *Neuropsychologia*, *50*, 2887–2896, doi:10.1016/j. neuropsychologia.2012.08.015.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Rovamo, J., & Virsu, V. (1979). An estimation and application of the human cortical magnification factor. *Experimental Brain Research*, *37*, 495–510.

Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General*, *112*, 585–615, doi:10.1037/0096-3445.112.4.585.

Streuber, S., Knoblich, G., Sebanz, N., Bülthoff, H. H., & de la Rosa, S. (2011). The effect of social context on the use of visual information. *Experimental Brain Research*, *214*, 273–284, doi:10.1007/ s00221-011-2830-9.

Theusner, S., de Lussanet, M., & Lappe, M. (2014). Action recognition by motion detection in posture space. *Journal of Neuroscience*, *34*, 909–921, doi:10. 1523/JNEUROSCI.2900-13.2014.

Thompson, B., Hansen, B. C., Hess, R. F., & Troje, N. F. (2007). Peripheral vision: Good for biological motion, bad for signal noise segregation? *Journal of Vision*, *7*(10):12, 1–7, doi:10.1167/7.10.12. [PubMed] [Article]

Thompson, J., & Parasuraman, R. (2012). Attention, biological motion, and action recognition. *NeuroImage*, *59*, 4–13, doi:10.1016/j.neuroimage.2011.05. 044.

Thornton, I. M., Wootton, Z., & Pedmanson, P. (2014). Matching biological motion at extreme distances. *Journal of Vision*, *14*(3):13, 1–18, doi:10. 1167/14.3.13. [PubMed] [Article]

Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., & Bülthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. *European Journal of Neuroscience*, *14*(5), 869–876, doi:10. 1046/j.0953-816x.2001.01717.x.

Tynan, P. D., & Sekuler, R. (1982). Motion processing in peripheral vision: reaction time and perceived velocity. *Vision Research*, *22*, 61–68, Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/ 7101752.