

Discrimination of locomotion direction in impoverished displays of walkers by macaque monkeys

Joris Vangeneugden

Laboratorium voor Neuro- en Psychofysiologie,
K.U. Leuven Medical School, Belgium



Kathleen Vancleef

Laboratorium voor Neuro- en Psychofysiologie,
K.U. Leuven Medical School, Belgium



Tobias Jaeggli

ETHZ Computer Vision Laboratory, ETH-Zentrum
Zürich, Switzerland



Luc VanGool

ETHZ Computer Vision Laboratory, ETH-Zentrum
Zürich, Switzerland, &
ESAT, K.U. Leuven, Belgium



Rufin Vogels

Laboratorium voor Neuro- en Psychofysiologie,
K.U. Leuven Medical School, Belgium



A vast literature exists on human biological motion perception in impoverished displays, e.g., point-light walkers. Less is known about the perception of impoverished biological motion displays in macaques. We trained 3 macaques in the discrimination of facing direction (left versus right) and forward versus backward walking using motion-capture-based locomotion displays (treadmill walking) in which the body features were represented by cylinder-like primitives. The displays did not contain translatory motion. Discriminating forward versus backward locomotion requires motion information while the facing-direction/view task can be solved using motion and/or form. All monkeys required lengthy training to learn the forward–backward task, while the view task was learned more quickly. Once acquired, the discriminations were specific to walking and stimulus format but generalized across actors. Although the view task could be solved using form cues, there was a small impact of motion. Performance in the forward–backward task was highly susceptible to degradations of spatiotemporal stimulus coherence and motion information. These results indicate that rhesus monkeys require extensive training in order to use the intrinsic motion cues related to forward versus backward locomotion and imply that extrapolation of observations concerning human perception of impoverished biological motion displays onto monkey perception needs to be made cautiously.

Keywords: biological motion, electrophysiology, object recognition

Citation: Vangeneugden, J., Vancleef, K., Jaeggli, T., VanGool, L., & Vogels, R. (2010). Discrimination of locomotion direction in impoverished displays of walkers by macaque monkeys. *Journal of Vision*, 10(4):22, 1–19, <http://journalofvision.org/10/4/22/>, doi:10.1167/10.4.22.

Introduction

Although recognizing actions of con- and heterospecifics is of utmost importance for an animal, we are still far from understanding how the brain accomplishes this feat. Current models of action recognition posit that the required computations are either based on an analysis of motion signals (e.g., Casile & Giese, 2005; Chang & Troje, 2009; Jhuang, Serre, Wolf, & Poggio, 2007; Thurman & Grossman, 2008; Troje & Westhoff, 2006), analysis of form signals (Beintema, Georg, & Lappe, 2006; Beintema & Lappe, 2002; Lange & Lappe, 2006), or both (Giese & Poggio, 2003; Schindler & Van Gool, 2008). Furthermore, the relative contributions of motion and form cues can be influenced by the task (Thirkettle, Benton, & Scott-Samuel, 2009).

Ultimately, an understanding of the neural basis of action recognition will require knowledge of how single neurons encode such stimuli. Previous research in awake, behaving monkeys has suggested that neurons in the macaque superior temporal sulcus respond selectively to action displays (Oram & Perrett, 1994a, 1994b, 1996; Vangeneugden, Pollick, & Vogels, 2009). However, despite the large volume of literature on biological motion recognition in humans (for reviews, see, e.g., Blake & Shiffrar, 2007; Giese & Poggio, 2003; Grossman, 2005), little is known about how non-human primates actually perceive biological motion. Previous electrophysiological work implicitly assumed that monkeys perceive actions in a manner similar to humans. Indeed, ethological observations by Wood, Glynn, and Hauser (2007) support this assumption. However, attempts to demonstrate the perception of

biological motion point-light displays in non-human primates have thus far been quite unsuccessful (Parron, Deruelle, & Fagot, 2007; Tomonaga, 2001), which may suggest that non-human primates' perception of some action displays might differ from that of humans.

In the present study, we trained rhesus monkeys to categorize the direction of locomotion of a “humanoid” walker. The walkers were visualized using cylinder-like geometrical primitives connecting the joints (Figure 1), and thus were richer in information than the point-light displays used in many human studies of biological motion but were reduced compared to real human walker displays. The movies were based on motion-capture data taken from real humans walking on a treadmill. Using stationary walkers instead of walkers who transverse a space eliminates translatory motion of the body through space, which is a strong but trivial cue for discriminating walking direction. Thus no extrinsic motion was present in the displays, only intrinsic (Chang & Troje, 2009). We trained the animals in two categorizations: (1) walking to the left by a leftward facing walker versus walking to the right by a rightward facing walker (view task) and (2) walking backward versus walking forward by a rightward facing walker (forward–backward task). The backward movies were reversed versions of the forward movies. Both tasks using these displays are accomplished effortlessly and with 100% accuracy by human observers.

The two tasks differ in the sort of information that distinguishes the categories. The view task can be solved using both motion and/or form cues, since both the motion trajectories and the postures of a leftward- and rightward-facing forward walker differ. However, the forward–backward task can be solved using motion information only since the postures are identical, i.e., that of the

rightward-facing walker. Mere form cues, without any temporal sequence analysis, cannot be used to distinguish forward from backward walking in these displays because they differed only in the temporal order of the snapshots.

In a first phase, we trained three monkeys in the two categorization tasks using displays of one actor walking at an intermediate speed of 4.2 km/h. To our surprise, we found that the 3 monkeys needed an extremely long training period to categorize displays of forward versus backward walking, displays that were immediately and effortlessly distinguished by human observers. After the training, we ran a series of generalization tests to determine the specificity of the categorization with regard to the trained displays. Thus, we tested for generalization to novel displays of the same actor walking and running at other than the trained speed, to different actors walking at the trained speed, to different stimulus formats including point-light displays, and to informative and non-informative parts of the walker. In addition, we parametrically increased the difficulty of the two tasks by manipulating the degree of spatiotemporal coherence in the locomotion displays. Finally, we examined the contribution of motion cues to the two types of tasks by degrading motion information.

General methods

Subjects and apparatus

Three rhesus monkeys (*Macaca mulatta*) served as subjects in this study: two females (M1: 6 kg; M2: 6.6 kg) and one male (M3: 7 kg). M3 participated in a previous

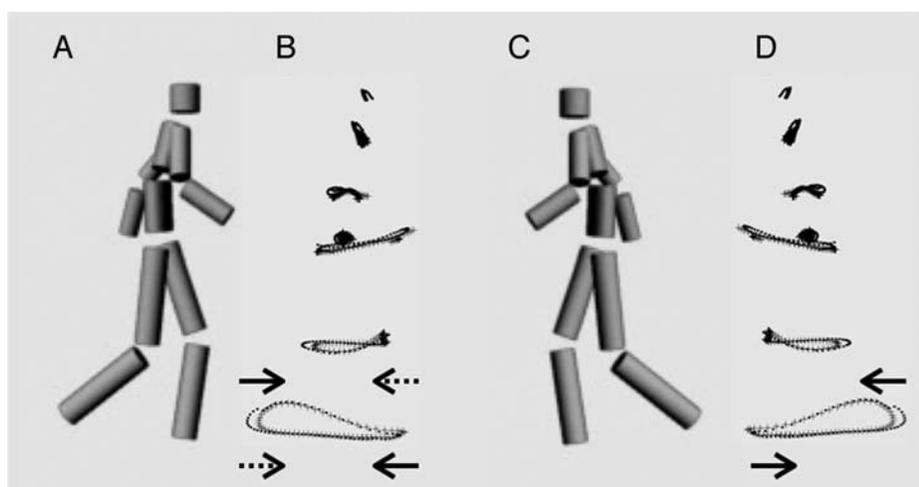


Figure 1. Snapshots of stimuli and motion trajectories. Single snapshots of a “humanoid” stimulus facing to the right (Rforw and Rback conditions) and to the left (Lforw condition) are shown in (A) and (C), respectively. The motion trajectories of the 15 major anatomical landmarks are shown for a single locomotion cycle in (B) and (D). Each marker indicates the position of the joint in a single frame. Filled and stippled arrows indicate the direction of motion of the ankle joints for forward and backward locomotions, respectively (see also movie: “[movie_fig1_categories.mov](#)”).

single-cell recording study using displays of stick figures performing various actions (Vangeneugden et al., 2009). The actions in that study consisted of transitive arm actions (such as throwing, knocking, and lifting); the animal was not trained to discriminate these actions but merely performed a simple fixation task during stimulus presentation. Thus, the present study was the first in which this animal was trained to actively discriminate stimuli, as was also the case for the other 2 naive animals. All monkeys had been trained in the fixation of a small target for juice rewards previous to the present study.

Prior to the experiments, a plastic head-fixation post was attached to the skull under aseptic conditions and isoflurane anesthesia. During the training and testing period, the animals were on a controlled fluid-deprivation schedule while dry food was available ad libitum in the home cage.

The position of one eye was monitored online via pupil position using an infrared eye tracking system (SR Research Eyelink; sampling rate of 1000 Hz). The stimuli were shown on a display, with a frame rate of 60 Hz, positioned at a distance of 57 cm from the animals' eyes. Stimulus presentation, timing, and juice delivery were controlled by a custom DSP-based computer system. This system also sampled the analog eye position signals from the tracker, checked eye fixation, saved and displayed online behavioral data. Tasks were controlled using custom-written software. During training and testing, the animals were seated in a primate chair with the head fixed.

All animal care, experimental and surgical protocols complied with national and European guidelines and were approved by the K.U. Leuven Ethical Committee for animal experiments.

Stimuli

The stimuli were generated using motion-capture data from 6 male human adults (age ranging between 20 and 40 years) of average physical constitution who were walking or running at various speeds on a treadmill. Walking speeds were 2.5, 4.2, and 6 km/h while running speeds were 8, 10, and 12 km/h. Unless otherwise stated, all movie renderings were performed at the average walking speed of 4.2 km/h for one actor; for clarity, we have designated these the “standard locomotion” conditions. The data were recorded at the Motion Capture Laboratory of ETHZ (Zürich) using an optical MoCap system (VICON) with 6 cameras operating at 120 Hz and a spatial resolution of 1 cm. In order to reconstruct the 3D body motions, subjects wore a skintight suit with 41 infrared-reflective markers placed on the major anatomical landmarks. The trajectories of the individual markers were then tracked and integrated into a 3D body representation. Based on these motion-capture coordinates, different displays were constructed using commercially available animation software (Maya, Autodesk, USA) or Matlab (The Mathworks, USA) for each speed

and actor. In the standard locomotion conditions, the displays consisted of humanoid figures where body limbs were represented by cylindrical geometrical primitives (Figure 1). Unless stated otherwise, the cylinders were shaded. The stimuli were presented at the center of the monitor against a light gray background. The height and width of the stimuli measured approximately 6 and 2.8 degrees of visual angle (maximum lateral extension of the ankles for the standard locomotion), respectively.

In all displays, the actors were viewed from the side (sagittal view) with locomotion from left to right or vice versa. From the 10-s-long motion-captured movies, segments of 1000 ms or, in later sessions, 1086 ms were extracted, approximating one full walking cycle for the standard walking speed of 4.2 km/h. The starting positions of the locomotion cycle were varied across the different movies by sampling up to 109 different segments from the full 10-s movie.

Tasks

The three monkeys had to perform 2 two-alternative categorization tasks: a view task and a forward–backward task. In the view task, the animals were required to categorize the facing direction of human locomotion, i.e., walking to the right (Rforw) versus to the left (Lforw). In the forward–backward task, they were required to determine whether a rightward facing actor was walking forward (Rforw) or backward (Rback). The Lforw displays were constructed by concatenating mirrored frames of the Rforw displays, while the Rback movies were created by reversing the frame sequence of the Rforw movies.

Each trial followed the same procedure. A trial started with the onset of a small red fixation target (size: $0.12^\circ \times 0.12^\circ$) that was presented at the center of the monitor, which the animal was required to fixate. After a fixation period of 500 ms, the locomotion was shown with the fixation target superimposed. The monkey was required to keep its gaze within a small fixation window (window size across monkeys: 1.3° – 1.7°) during the entire fixation and stimulus period. Failure to do so was penalized by removing the fixation target and stimulus. After presentation of the stimulus, the fixation target was replaced by 2 target squares (size: $0.4^\circ \times 0.4^\circ$). Stimulus categories were associated with different target square positions: for the Rforw and Rback categories, the targets were located along the horizontal meridian 8.4° to the right and the left of the center of the screen, respectively, while the target for the Lforw category was 8.4° up along the vertical meridian. An immediate saccade to the correct target was rewarded with a drop of apple juice. Eye positions outside the predefined window during the fixation or stimulus period as well as saccades outside the target windows were considered as aborts and were not incorporated in the data analyses. Thus, only trials in which the saccade landed in one of the target windows were analyzed. The

target windows were on purpose rather large (approximate size: $5^\circ \times 5^\circ$) and the monkeys could easily saccade toward the target points.

Locomotion categorization training

Left- and rightward facing walkers differ in the motion trajectories of their limbs and joints (Figures 1A, rightward and 1C, leftward) as well as with regard to spatial and form cues such as the curvature of the back and orientation of the knee angles. Thus successful performance in the view task, i.e., discriminating the facing direction of locomotion, can rely on motion and form cues though form information alone should be sufficient to solve the task (Beintema et al., 2006; Beintema & Lappe, 2002). However, the Rforw and Rback stimuli in the forward–backward task consist of the same snapshots (i.e., frames) and differ only in the direction of the motion along the trajectories or the temporal sequences of the snapshots. The difference between the Rforw and Rback motions of the joints is indicated for the ankle joint by the solid and stippled arrows in Figure 1B. The spatial patterns of the trajectories are identical for forward and backward walking, only the direction of motion differs. For the other joints, the difference in the motion patterns are smaller and nearly absent above the knees. Thus, the successful discrimination of forward versus backward locomotion cannot be based on spatial cues only but requires at least some motion analysis, i.e., an analysis of the local motion trajectories per se, or a temporal integration of body poses (Lange & Lappe, 2006). Thus far, there have been no indications as to how macaque monkeys might perform on these two types of locomotion tasks, particularly so because no extrinsic translatory motion cues are present. Therefore, we trained three macaque monkeys in both the view and the forward–backward task.

Stimuli

We used the standard, shaded humanoid figures (Figures 1A and 1C) based on motion-capture data from one actor walking at 4.2 km/h.

Tasks

All three monkeys participated in the 2 two-alternative categorization tasks. The order of training of the categorization tasks was counterbalanced across animals. M1 was first trained in the view task and then in the forward–backward task, while the opposite order was followed with the two other animals.

Training procedure

Training in each task started with trials in which only the correct target square was presented after stimulus presentation. Consequently, saccades into the target windows of these locations were rewarded. In this manner, the monkey learned to associate the correct target with a particular locomotion category. Then, trials with two target squares (double target trials) were intermixed with trials with only one target while the proportion of the latter sort of trials was gradually reduced. In monkeys M2 and M3, we also employed a response-bias-correction procedure in some training sessions since both subjects had strong tendencies to respond to just one particular target irrespective of the locomotion category. In this bias-correction procedure, a trial in which an error occurred was followed by a trial in which the same stimulus category was shown as that in the error trial. This procedure was maintained until a correct response was made. Blocks with bias correction were interleaved with shorter blocks (approximately 150 trials) of regular trials without bias correction. Daily training sessions contained approximately 1100 trials, with some intersubject variability.

Data analysis

The performances of the 3 monkeys were analyzed separately. Training curves were computed using the proportion of correct responses in the trials in which the two targets were presented simultaneously, averaged across the two alternatives, for each daily session.

Results and discussion

Figure 2 shows the training curves for the 3 monkeys. It is obvious that each of them required many more training sessions to learn the forward–backward compared to the view task. Only a couple of sessions, at most, were necessary to become proficient in the view task while an extremely large number of training sessions were required to learn the forward–backward discriminations. Moreover, to train the latter task, we needed either a large number of single-target training blocks, as in monkeys M1 and M3 (indicated by red symbols in Figure 2) or a bias-correction procedure as in monkeys M2 and M3 (indicated in green in Figure 2).

Note that the percentages of correct responses shown in Figure 2 are computed only for trials in which two target points were shown simultaneously and for blocks without bias correction. The number of trials required to reach 75% correct (in a session) for the view task was 5417 (4 sessions), 1323 (1 session), and 2272 (2 sessions) for M1, M2, and M3, respectively. Many more trials were required to attain 75% correct for the forward–backward task: 18,023 (19 sessions), 37,238 (39 sessions), and

43,576 (29 sessions) for M1, M2, and M3, respectively. Note that the profound difference between the two tasks in each animal cannot be explained by the order of training, since the latter differed among animals.

The view task can be solved using spatial or motion cues, while successful performance in the forward–backward task requires the use of motion information. The present

data show that for locomotion displays that contain no extrinsic motion cues—because of the lack of overall translatory motion—the motion information that distinguishes forward from backward has a low perceptual saliency. This is in striking contrast to humans who easily and effortlessly discriminate the forward- and backward-walking displays used here.

Stimulus specificity of the acquired locomotion categorizations

After achieving proficiency in the two-alternative categorization tasks, the monkeys were tested for their ability to generalize to other locomotion displays. Thus we could determine how specific the acquired categorization was for various aspects of the stimulus displays and locomotion mode. We tested generalization of walking and running speeds, different actors, stimulus impoverishment such as point-light displays and the presentation of only parts of the body.

Stimuli

The following stimuli were employed in the generalization tests:

(1) Walking/running speed variations: humanoid displays of the same actor locomoting at different speeds (2.5, 6, 8, 10, and 12 km/h; standard speed = 4.2 km/h). These were based on actual motion-capture data of the same actor (Figure 3A, lower panel).

(2) Actor variations: humanoid displays of 5 different actors motion-captured while walking at 4.2 km/h (Figure 3B, lower panel).

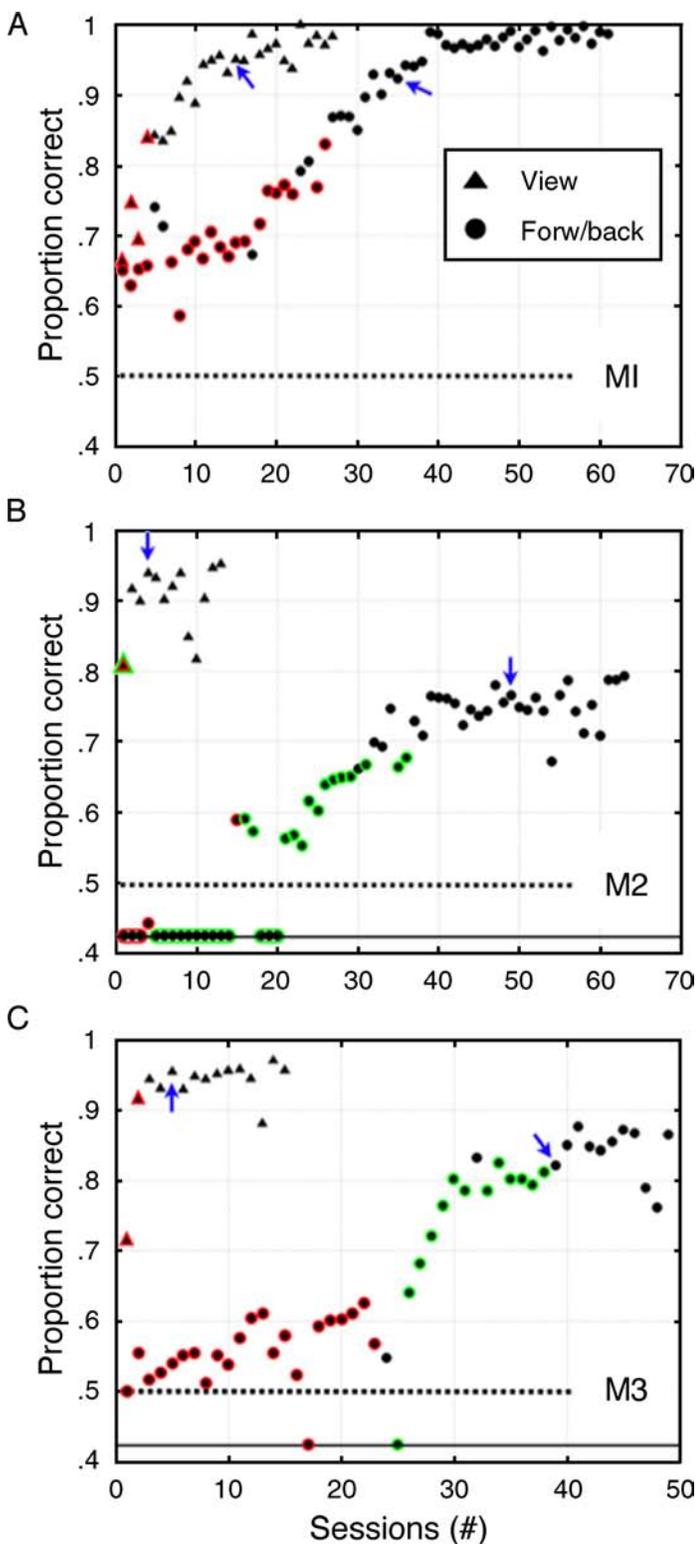


Figure 2. Learning curves for the view and forward–backward categorization tasks. Proportion of correct responses is plotted as a function of training session for the view (triangles) and forward–backward (circles) tasks for each monkey separately (A, M1; B, M2; and C, M3). On average, each session contained approximately 1100 trials. Chance level (50%) is indicated by the dotted line. Symbols in red and green on the bottom solid horizontal line represent sessions in which only single-target or only bias-correction trials were shown, respectively, and thus for which no performance measures were available. Performance on regular double-target trials, in blocks interleaved with bias-correction blocks, is indicated by green symbols, performance on regular double-target trials interleaved with single-target trials is indicated by red symbols, and performance in sessions consisting of only double-target trials is indicated by black symbols. Blue arrows mark the start of the generalization testing procedure for each task separately.

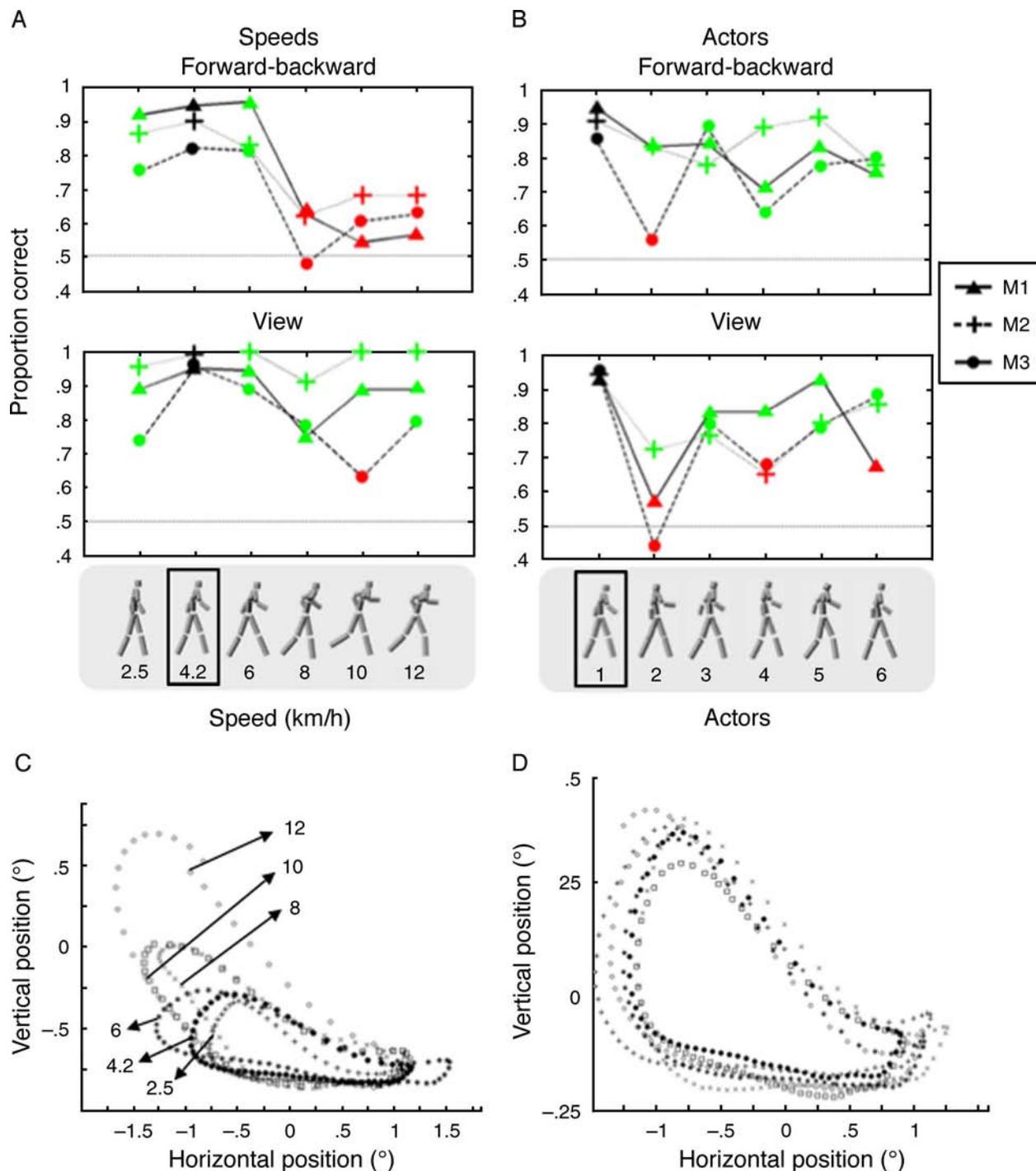


Figure 3. Generalization for speed and actor. Proportion of correct responses plotted as a function of speed (in kilometers per hour; A) and for 6 different actors (B) for each of the 3 animals (M1, M2, and M3). The trained speed (4.2 km/h) and actor (#1) are indicated by black boxes on the corresponding stimulus panels showing snapshots from the locomotion movies. Performances are plotted separately for the forward-backward (upper panel) and view task (lower panel). Green markers indicate performances significantly different from chance level (50%, indicated by a horizontal line; binomial test: $p < 0.05$) while red markers indicate performances that did not differ significantly from chance. The black markers represent the performance on the trained standard locomotions. Performance for the generalization trials are calculated for an average of 20 to 30 trials for each condition separately, while the data for the trained standard locomotions presented in the same test are based on approximately 1000 to 1500 trials. The motion trajectories of the ankle joint are plotted for the different speeds and actors in (C) and (D), respectively. The different walking and running trajectories are indicated by arrows (C) while trajectories for the different actors, due to their high degree of similarity, are left unassigned. Filled circles represent the trained locomotion (see also movies: "movie_fig3_speeds.mov" and "movie_fig3_actors.mov").

(3) Manipulations of the stimulus format: (a) Point-light displays: 15 black dots, each measuring 0.43° in diameter, replaced major anatomical landmarks, i.e., joints and head (Figure 4C, left panel). The same actor and walking speed as those of the standard locomotion condition were used. (b) Stick-plus-point-light displays: the above-mentioned 15 points were connected with thick lines (Figure 4C, middle panel). Again, the same actor and walking speed as those of the standard locomotion condition were used. (c) “Puppet” displays: the 4.2 km/h walker was rendered as a clothed human male. This stimulus was achromatic (Figure 4C, right panel).

(4) Body part displays: upper and lower body parts of the shaded humanoids, i.e., above and below the hip, respectively, were presented in isolation (Figure 5C).

Generalization test

In these generalization tests, the trained humanoid locomotion, i.e., actor 1 walking at 4.2 km/h, was shown in about 90% of the trials, while the remaining 10% were so-called generalization trials in which the novel generalization stimuli were presented. For each generalization stimulus category tested (e.g., a different speed), we presented different movies (usually $n = 5$ for each stimulus category) by varying the starting position within the full walking/running cycle. Since we wished to measure the spontaneous categorization behavior of the animal in the generalization trials and thus not train him or her merely to respond to the correct target, saccades toward either of the two target locations were rewarded in these generalization trials (see Vogels, 1999). In the interleaved, and more frequent, regular trials (presenting the trained standard humanoid locomotion), only correct responses were rewarded. Most generalization tests contained between 1000 and 1500 trials, thus presenting approximately 900–1350 regular versus approximately 100–150 generalization trials.

Data analysis

The proportion of generalization trials in which the monkey responded to the correct learned-category targets, averaged across the two alternatives, was taken as measure of generalization. To test whether the generalization was significantly different from random choice (50% correct), we used a binomial test (Vogels, 1999).

Results and discussion

Speed variations

Figure 3A shows the generalization performance for the 3 monkeys for variations in speed. The data of M2 were obtained after prolonged training involving the three-alternative task (>50,000 trials; see [Degrading](#)

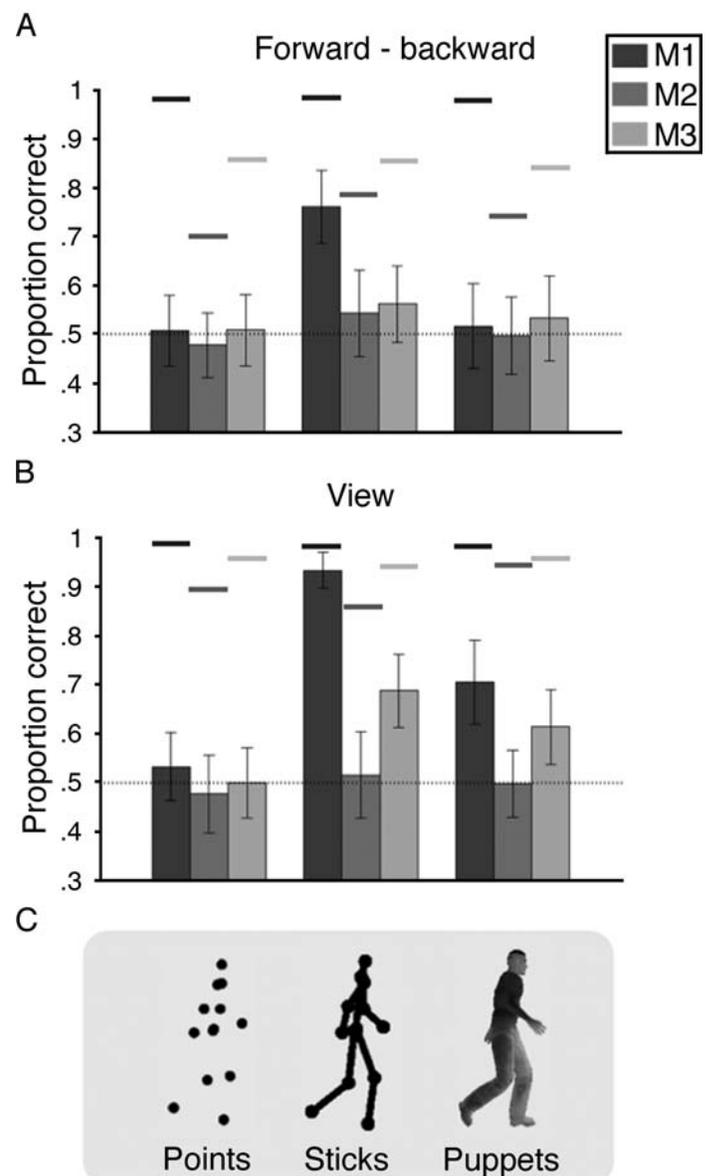


Figure 4. Generalization performance to different stimulus formats. Snapshots of the movies are shown in (C): point-light, stick-plus-point-light, and anthropomorphic puppet displays. Performances are plotted for the (A) forward–backward and (B) view task for each monkey (M1, M2, and M3) separately. Error bars indicate 95% confidence intervals while chance level is indicated by the dotted line. The floating bars above each histogram plot performance for the trained locomotions, measured in the same tests. Data for the generalization trials are calculated for an average of 100 to 150 trials for each condition separately, while the data for the trained standard locomotions presented in the same test are based on approximately 1000 to 1500 trials (see also movie: “[movie_fig4_stimformats.mov](#)”).

[spatiotemporal coherence of the walker](#) section). This explains her relatively good performance for the trained speed, especially in the forward–backward task (approximately 90% correct), compared to her performance in the

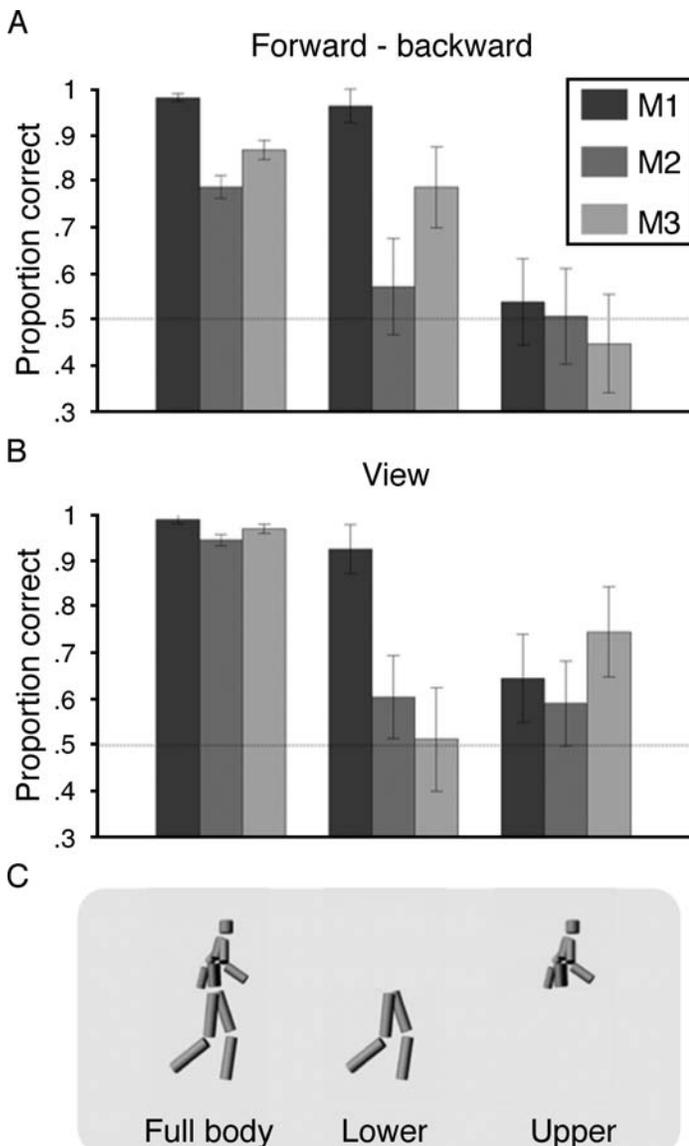


Figure 5. Generalization performance for the lower and upper body parts of the humanoid locomotions. (A, B) Performances for the forward–backward and view task for each monkey (M1, M2, and M3), respectively. (Left) Performance for the trained full-body movies, (middle) generalization performance for the lower body displays, (right) generalization performance for the upper body displays. (C) Snapshots of the displays. Same conventions as in Figure 4. Generalization performance was calculated on approximately 40 trials per body configuration (i.e., lower or upper) per condition (see also movie: “movie_fig5_parts.mov”).

final stages of the initial categorization training (approximately 80% correct in the forward–backward task). It also shows that after extensive training, this animal reached a performance level in the forward–backward task similar to that of M1. The speed generalization tests of M1 and M3 were performed after the initial training phase, thus starting in the session indicated by the blue arrows in Figure 2. The data from the speed-generalization tests

showed that the forward–backward categorization was specific to walking: in each monkey, generalization was significant (binomial test; $p < 0.05$: green markers in Figure 3A) for the walking, but not the running patterns. In fact, in each monkey there was an abrupt drop of the performance when the locomotion changed from walking to running. The lack of significant transfer from the trained walking to running suggests that the animals learned a particular motion trajectory “template” in the forward–backward task. Indeed, examination of the ankle trajectories (Figure 3C) reveals a relatively high similarity between those trajectories for the three walking speeds (2.5, 4.2, and 6 km/h), which are in turn rather distinct from those of the three running patterns (8, 10, and 12 km/h). In contrast to the forward–backward task, categorization in the view task generalized relatively well across the different walking and running speeds. This suggests that the discrimination in the view task is based on spatial or motion cues that are common to the different speeds. The broader generalization observed in the view task compared to the forward–backward task shows that such motion cues are less specific in the former. Alternatively, the monkeys might have used spatial features that are common to the walking and running humanoids that face in a particular direction.

Actor variations

Overall, generalization to different actors was evident across tasks and monkeys, except with actor 2 for monkeys M1 and M3 in the view task and monkey M3 in the forward–backward task (Figure 3B, the forward–backward generalization data for M2 were obtained after prolonged training with the three-alternative task; see [Degrading spatiotemporal coherence of the walker](#) section). Performance in the view task with actor 4 for M2 and M3 and for actor 6 for M1 was just marginally non-significant. The motion trajectories of all actors are highly similar (Figure 3D) suggesting the contribution of form cues to the discrimination. More specifically, the lack of transfer, especially for actor 2 in the view task, could be the result of the somewhat different overall upper body configuration of this actor (a shorter neck and straighter back) compared to the other actors (Figure 3B, lower panel). The actor-generalization results suggest that for both types of categorization there is some influence from the spatial or body configuration.

Manipulations of the stimulus format

How specific were the learned categorizations to the humanoid figures? This was tested by presenting point-light displays, stick-plus-point-light figures, and anthropomorphic puppets (Figure 4C) that were based on the joint trajectories calculated from the same motion-captured data that we used to generate the humanoids. No generalization was present in either task for point-light displays of the trained actor performing the trained locomotion: for each task

and animal, performance was close to 50%, i.e., chance (Figures 4A and 4B). Moreover, monkey M2 also demonstrated a lack of transfer to the stick-plus-point-light and puppet displays, while monkeys M1 and M3 were able to perform significantly above chance level (binomial test; $p < 0.05$) for both these displays, at least in the view task. For the forward–backward task, significant generalization was present only for the stick-plus-point-light displays but only for M1 (binomial test; $p < 0.05$). Note that even for M1, who showed generalization in the view task for the puppets and stick-plus-point-light figures, her generalization for the point-light displays was at chance level. This indicates that stimuli depicting only the joint trajectories were insufficient to support the categorizations.

Body part displays

Figure 1B shows that differences in direction of the motion trajectories are present for the legs (ankle and knee joints) but nearly absent in the upper body parts. Thus, one would expect that discriminating between forward and backward categories would not be possible when only the upper part of the walker is presented, whereas it might generalize to a display of only the legs. Generalization to the lower body parts will depend on whether the monkeys' categorization performance is affected by the presence of the whole-body configuration. Therefore, we tested the generalization performance for locomotion displays of the upper and lower body parts of the trained actor. As shown in Figure 5, generalization performance for the forward–backward task hovered around 50% in each of the 3 animals when only the upper body was presented. Performance for the lower body was well above chance for monkeys M1 and M3 but not for M2. In fact, generalization to the lower limb displays was almost as high as the performance for the trained locomotions in animals M1 and M3. The lack of transfer for the lower limbs in monkey M2 indicates that this monkey was sensitive to the presentation of the whole body configuration despite the fact that the informative cues were present only in the bottom part. Generalization to the part displays in the view task indicated that both top and bottom limbs were used in this task, with the contribution of each part depending on the monkey. This is in line with the fact that both upper and lower limbs contain form and motion cues that can distinguish leftward versus rightward walkers, and different animals may use differently located cues.

Degrading spatiotemporal coherence of the walker

Next, we designed a dynamic scrambling procedure to parametrically impair the spatiotemporal coherence of the

walker in the two tasks. We varied the degree of coherence by randomly repositioning a given proportion of small divisions (i.e., tiles) of the walker image. The degree of walker coherence was defined by the proportion of parts that were not scrambled but remained at their original and thus correct position. This manipulation was designed to affect both spatial and temporal information: spatial information was degraded by the (partial) scrambling, while temporal information was disrupted by repeating this scrambling operation after a certain number of frames (i.e., every frame or every fifth frame). This additional manipulation impairs the motion trajectories of the limbs. We therefore expected that the spatiotemporal scrambling procedure would have an effect on both view and forward–backward discriminations. Furthermore, we expected that the forward–backward discriminations would be more strongly impaired than the view task discriminations, i.e., forward–backward discriminations would tolerate less incoherence than view discriminations. The latter could still be solved by integrating the posture information across frames, even at low coherence levels, while the former crucially depends on the motion direction of the lower limbs, which are strongly affected by the dynamic, spatiotemporal coherence manipulation.

To directly compare the computation of view and the forward–backward locomotion mode, we trained the animals in a three-alternative categorization task in which three target squares were presented. The monkeys were thus obliged to simultaneously discriminate between Rforw, Rback, and Lforw categories. Solving this three-alternative task requires both a decision about the view and a decision about the forward–backward mode of locomotion, at least in the case of a rightward facing walker. Note that discrimination of overall walking direction is insufficient to solve this task, since walking direction is the same for the Lforw and Rback stimuli. By analyzing the confusion matrices of the behavioral responses in this task (see Data analysis section), we can determine what sort of information (overall walking direction, view and/or forward versus backward mode) is used by the animal and how this is affected by the spatiotemporal coherence.

Stimuli

We employed different coherence levels ranging from 100% (unscrambled) to 0% (fully scrambled) in steps of 10% (Figure 6C; coherence levels in steps of 20%). First, we defined a region (Figure 6A; mask canvas) that encompassed the humanoid extremities for all walking locomotions (2.5 to 6 km/h). This region was divided into square tiles measuring 8 by 8 pixels (i.e., 0.23°). The scrambling procedure itself was performed by randomly repositioning the tiles within the larger region of the mask canvas. The degree of coherence was manipulated by repositioning a fixed proportion of the tiles while leaving the other tiles at their original position. Thus, in a 50%

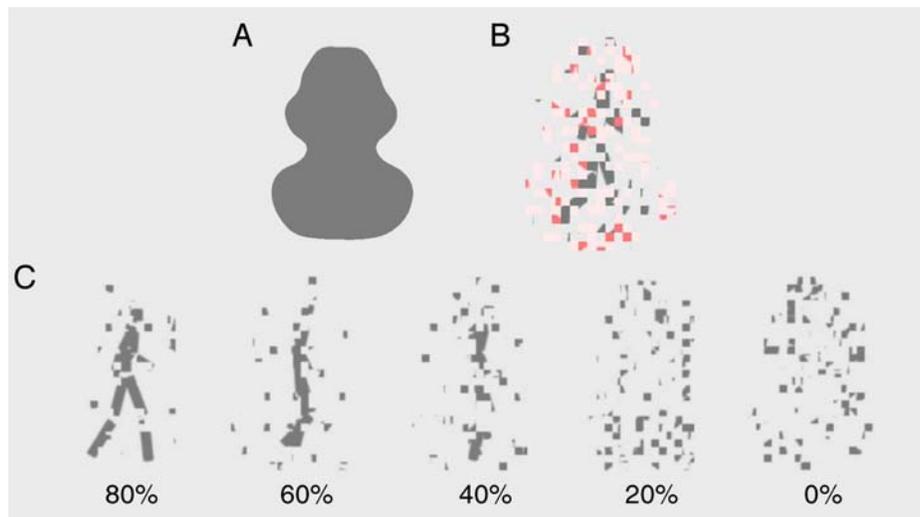


Figure 6. Spatiotemporal coherence manipulation. (A) The mask canvas, generated by smoothing the superimposed arrangement of all frames from the three walking speeds, defined the area in which tiles could be randomly reallocated. (B) Repositioned tiles consisted of 50% tiles from the original facing direction (e.g., Rforw: gray tiles in (B); shown for a 40% coherence stimulus) and 50% tiles from the opponent facing direction (e.g., Lforw: red tiles in (B)). (C) Single frames randomly extracted along the walking cycle for 5 spatiotemporal coherence levels (see also movie: “[movie_fig6_scrambling.mov](#)”).

scrambled locomotion, 50% of the tiles were repositioned while the other 50% remained at their original location. The orientations and shading patterns of the cylinder fragments in the tiles can differ between the Rforw and Lforw conditions. These low-level cues might therefore be used to discriminate between scrambled displays of these stimuli. To avoid the use of such cues, we removed the shading component from the cylindrical primitives, i.e., an unshaded humanoid, and we insured that half of the repositioned tiles were from the scrambled stimulus (e.g., Rforw) while the other half were from the complementary, mirrored frame depicting an actor facing the other direction (e.g., Lforw; see [Figure 6B](#)). Thus, for an Rforw display with a scrambling level of 50%, 25% of the scrambled tiles originated from the Rforw stimulus and the other 25% of the scrambled tiles came from a corresponding frame of the Lforw stimulus. Monkeys M1 and M2 (M3 was not involved in this part of the study) generalized significantly to the unshaded humanoids with average performances across the 3 categories for M1 and M2 of 72.6% and 57.1%, respectively (chance performance = 33.3%; binomial tests for all 3 combinations of two categories; $p < 0.05$).

In the 5-frame lifetime condition, the repositioning of the tiles was repeated every 5th frame (i.e., 83 ms). Thus, small parts of the walking humanoid were visible for 5 frames at the correct location surrounded by randomly placed segments consisting of the cut-out portions of the same humanoid stimulus either walking in the same or opposite directions (mirror image of the cut-out tile), both having 50% chance of occurrence. More specifically, the procedure for the coherence manipulation was as follows: superimpose a grid over the complete mask canvas, cut out

a specific number of small tiles and randomly position them within the mask canvas for a duration of 5 frames. In the 1-frame lifetime condition, the scrambling operation was performed every frame, impairing local motion information even more strongly. For each scrambling level and for each of the 3 categories (Rforw, Rback, and Lforw), 14 movies with different starting positions of the walking cycle were presented. The starting positions covered the complete walking cycle.

Task

The three-alternative categorization task follows exactly the same procedure as the two-alternative tasks, except that the three categories are shown randomly interleaved in a test and thus three instead of two target points are shown simultaneously. Monkeys M1 and M2 were successfully trained in this task. During training, the coherence level was gradually decreased. After extensive training for about 2 months using the 5-frame lifetime conditions, we collected data from M1 and M2 with both the 1- and 5-frame conditions for the three categories in the same test. For both lifetimes, 6 different coherence levels (100%, 80%, 60%, 40%, 20%, and 0%) were presented in an interleaved fashion with approximately 145 and 140 trials per condition per lifetime for M1 and M2, respectively.

Data analysis

The data from the three-alternative task was analyzed in different ways. First, we computed the proportion of

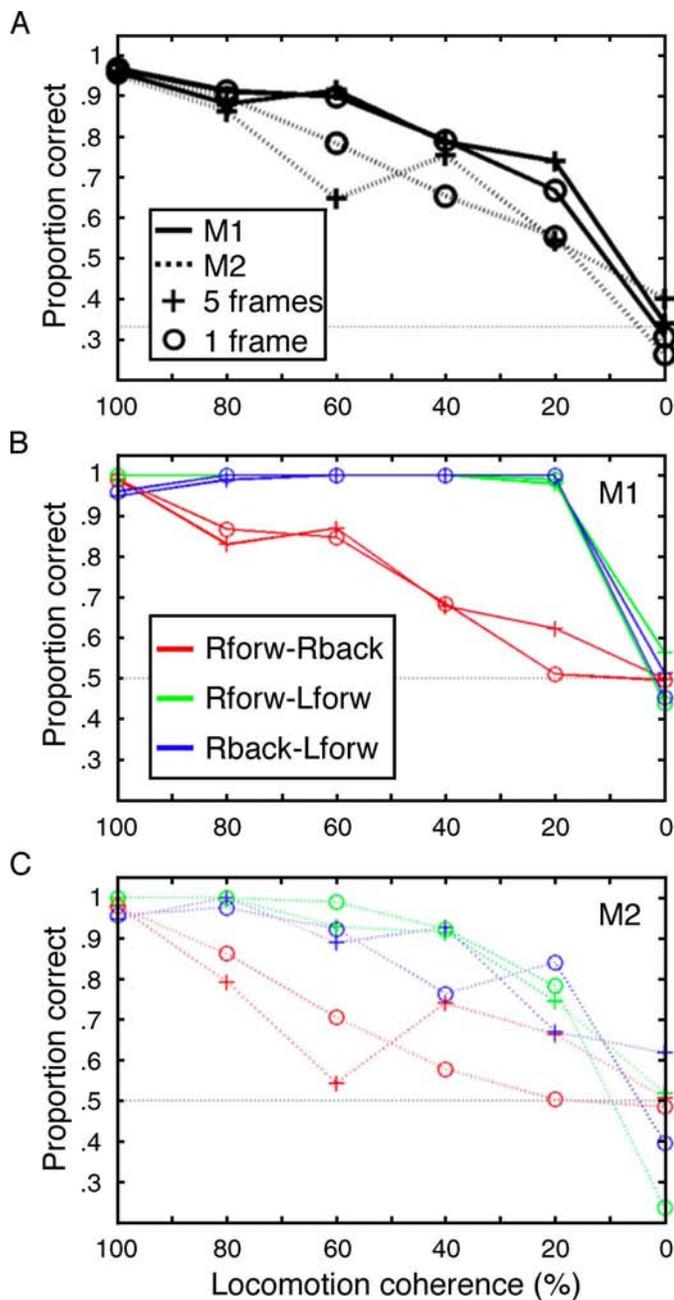


Figure 7. Effects of degrading the spatiotemporal coherence of the walker in the three-alternative discrimination task. (A) Psychometric curves plotting the average proportion correct as a function of decreasing stimulus coherence levels for M1 (full lines) and M2 (dotted lines) separately. The two different intervals before tile reshuffling reoccurred are indicated by crosses (5-frame lifetime) and circles (1-frame lifetime), respectively. Chance performance is equal to 0.33 correct. Both monkeys were approximately exposed to 140 trials per coherence level per lifetime. (B, C) Discrimination measures for each combination of two categories at the various locomotion coherence levels for (B) M1 and (C) M2, respectively. Chance performance is equal to 0.50. Different colors refer to the possible combinations of two categories: red = Rforw versus Rback, green = Rforw versus Lforw, and blue = Rback versus Lforw. Same conventions as in (A).

correct responses, averaged across the three alternatives. Second, we produced confusion matrices, tabulating the frequency of each of the three choices (i.e., responses) given each stimulus category. From these confusion matrices, we computed three discrimination measures: discrimination of Rforw versus Rback, discrimination of Rforw versus Lforw, and discrimination of Rback versus Lforw. In this last comparison, the general locomotion direction was the same (to the left), while these differed for the other two comparisons. These three discrimination measures were computed using the constant ratio rule (Clarke, 1957; Macmillan & Creelman, 2004). For instance, as a forward versus backward category discrimination measure we computed the percent correct responses based on the frequencies of Rforw and Rback target responses in those trials for which locomotion in one of these two categories was presented.

Results and discussion

As expected, both monkeys showed overall performance decreases with decreasing coherence for both the 1- and 5-frame lifetime conditions (Figure 7A). The rate of change of the noise had relatively little impact on the overall performance (compare performance for the 1- and 5-frame lifetime conditions; Figure 7A). Since three stimulus alternatives are presented in this task, one can distinguish between three possible discriminations. These discrimination measures are plotted in Figures 7B (data from M1) and 7C (data from M2) as a function of coherence level. The forward–backward discriminations were more strongly affected by the coherence manipulation than the other discriminations in which both form and motion cues were available. In both monkeys and for both lifetimes, all comparisons between the forward–backward and the other two discrimination measures showed a significant difference (binomial test; $p < 0.05$, two-sided) except for the 100% and 0% stimulus coherences and the 20% coherence level in M2 for the 5-frame lifetime (crosses in Figure 7C). In both animals, forward–backward discriminations were already impaired at the 80% coherence condition and were close to chance at the 20% coherence level for M1. Note that, even at this 20% level, both animals, especially M1, were easily able to discriminate locomotions when they differed in both motion and form. The stronger effect of spatiotemporal coherence on the forward–backward in comparison to the other discriminations is consistent with its overall greater difficulty, i.e., the much longer time it took to learn the forward–backward discriminations. The discrimination measures at the different coherence levels of Lforw versus Rforw were similar to those for the Lforw versus Rback conditions.

A likely reason that manipulating the spatiotemporal coherence had a stronger effect on the forward–backward discrimination performance is that the latter task can be

solved only by using motion information, which is degraded by the coherence manipulation, while the other discriminations can be solved using body-posture-related form cues, in addition to motion. In the 5-frame lifetime conditions, i.e., rescrambling the stimulus only after every 5th frame, motion information is less degraded than in the 1-frame lifetime conditions. Thus one would expect, in the forward–backward discrimination, an effect of frame lifetime on performance at low coherence levels. There was a trend for such an effect in monkey M1: 51% versus 63% correct (binomial test; $p < 0.05$: one-sided) for the 1- versus 5-frame condition, respectively, at 20% coherence. Monkey M2 showed more variable frame-lifetime effects on the forward–backward discrimination: higher performance on the 1- compared to the 5-frame condition at 60% coherence (70.5% versus 54.3%; binomial test; $p < 0.05$: one-sided) and the expected higher performance for the 5- compared to the 1-frame condition at the 40% and 20% coherence levels (74% versus 58% and 67% versus 50%; binomial tests; $p < 0.05$: one-sided).

Degrading motion information

The previous manipulations of spatiotemporal coherence affected both form and motion information. Next, we wished to manipulate motion information without affecting form information. Following previous studies of the contribution of motion information to the perception of biological motion in point-light displays (Beintema et al., 2006; Mather, Radford, & West, 1992; Thornton, Pinto, & Shiffrar, 1998), we degraded motion information by using displays in which a fixed number of frames were replaced by blanks, keeping the apparent speed constant. Such semi-stroboscopic displays (“semi” since the duration of a single frame is longer than that typically used in real stroboscopic displays) are expected to affect the direction selectivity of early visual cortical motion areas such as monkey area MT, where direction selectivity has been shown to decrease with increasing stroboscopic interflash intervals (Churchland & Lisberger, 2001; Mikami, Newsome, & Wurtz, 1986; Newsome, Mikami, & Wurtz, 1986). The shortest interstimulus intervals (ISIs) at which direction selectivity is still present in MT neurons further depends on several factors such as the stimulus speed and preferred speed of the neuron (Churchland & Lisberger, 2001). Given that the speeds present in our locomotion stimuli are relatively slow and thus activate mainly neurons preferring slow speeds, one would expect direction selectivity to be affected when ISIs last longer than about 30 ms (Churchland & Lisberger, 2001). In addition, based on the relationship between preferred speed and effect of ISI on MT responses, the ISI may affect the speed estimated over a population of neurons (Churchland & Lisberger, 2001). Thus, if the locomotion direction discrimination depends

on short-range motion mechanisms (Braddick, 1974), as implemented by MT neurons, one would expect it to break down at ISIs above 30 ms.

In the ISI manipulations, longer ISIs are confounded with larger spatial displacements since the apparent speed was held constant. To assess the contribution of the longer spatial displacements that resulted from increasing ISIs, we tested an additional display in which the displacements were identical to those used in the ISI manipulation but differed from the latter by having a still frame for the entire ISI. These displays result in a jerky type of motion and thus we label these as “jerky” locomotion displays.

The final manipulation used to degrade motion was an extreme one: the presentation of a single static snapshot. This manipulation was tested only for the view task, since obviously such static snapshot displays cannot provide information regarding forward–backward locomotion. However, they contain form/posture information that should, at least in principle, be sufficient to solve the view task.

The three sorts of displays (ISI, jerky locomotion, and static displays) were tested in generalization tests (see [Stimulus specificity of the acquired locomotion categorizations](#) section) in order to measure the subjects’ ability to generalize from the standard locomotion conditions to these novel displays without extra training.

Stimuli

ISI manipulations

Snapshots were omitted from the temporal sequence and replaced by blank intervals. Thus, each visible frame was followed and preceded by a fixed number of empty frames, keeping the 4.2 km/h walking speed (e.g., a sequence for a 3-frame ISI: 1./././5./././9./...; ‘/’ refers to blank frames while the numbers indicate frames of the original sequence). The number of blank frames between visible frames was fixed for a particular sequence but varied between sequences (5 ISIs; range: 1 to 9 frames with a step size of 2 frames). Since we used a CRT display, the stimulus duration was shorter than the frame duration. Measurements with a photodiode showed that the stimulus on our CRT display, at a particular location, was displayed for approximately 2 ms at a rate of 60 Hz. Thus the effective ISIs ranged from approximately 30 ms (1-frame ISI) to approximately 164 ms (9-frame ISI).

Jerky locomotions

These stimuli were created by manipulating the snapshot duration while maintaining the standard walking speed of 4.2 km/h. The number of still frames was fixed within a sequence, e.g., a sequence with 3 additional still frames: 1,1,1,1,4,4,4,7..., but varied between sequences: range between 2 and 6 frames with a step size of 1 frame (33 to 100 ms). When displayed on our CRT display with

a frame rate of 60 Hz, perception of smooth motion was disrupted for the higher number of still frame conditions.

Static snapshots

In total 14 snapshots were extracted from the full locomotion cycles for each of the Rforw and Lforw conditions of the standard locomotion and these 28 snapshots were presented for either 300 or 1086 ms. The 14 snapshots spanned the entire range of the full locomotion cycle.

Tasks

Only M1 and M2 were tested. The stimulus conditions were presented according to the generalization protocol described previously and thus were interleaved with the original standard locomotions. The static snapshots were displayed only in the two-alternative view task. We collected data for approximately 18 and 10 generalization trials for each of the 14 static snapshot stimuli (combined for Rforw and Lforw conditions) for M1 and M2, respectively. The ISI and jerky displays were shown in both the three-alternative and 2 two-alternative tasks, with the exception of the jerky displays, which were presented only in the two-alternative direction task (i.e., forward versus backward discrimination). Thus, for the jerky displays we collected data for the full manipulated range of still frames (i.e., 1 to 5 additional still frames) in the forward–backward task, while we collected data only for the 3-still-frame sequence in the three-alternative task. For the ISI displays, we collected data for all manipulations in both two-alternative tasks.

In the jerky-locomotion tests, an average of 24 (M1) and 31 (M2), or 52 (M1) and 207 (M2) generalization trials were shown for each condition in the forward–backward and the three-alternative task, respectively. In the ISI tests, the number of generalization trials shown averaged 33 (M1) and 40 (M2) for each condition in the forward–backward task, and 23 (M1) and 19 (M2) for the view task. Note that in all these generalization tests the standard locomotions were shown with a tenfold frequency with respect to the sum of all generalization trials in that test.

Data analysis

The same kinds of data analyses were performed as in the other generalization tests (see above). We focused mainly on the data obtained in the two-alternative tasks, except for the jerky Rforw–Lforw discriminations since the latter were presented only in the three-alternative task. The computation of correct generalization responses in the

three-alternative task was based on application of the constant ratio rule, as explained above.

Results and discussion

Degrading motion information by prolonging the ISI had a dramatic effect on the performance in the forward–backward categorizations (Figures 8A and 8B; M1 and M2, respectively). Even introducing a single blank frame, i.e., an ISI of approximately 30 ms, reduced the forward–backward generalization abilities of both animals considerably. Monkeys M1 and M2 obtained 64% and 68% correct, respectively, for the 1-frame ISI compared to 99% and 97% correct for the standard locomotions, presented in the same test (binomial test; $p < 0.0001$, one-sided). The strong ISI effect was not due just to a disruption of the animals' behavior by the flicker in the ISI displays, since the direction categorization of the same flickering displays was much less strongly affected. Indeed, the performance averaged across the five ISIs was 82% and 81% correct for M1 and M2, respectively, in the view task, but only 52% and 60% correct for M1 and M2, respectively, in the forward–backward task. The difference in the proportion of correct generalizations was highly significant (binomial; $p < 0.0001$: data pooled across ISIs) in each animal. These data thus indicate that even a slight degradation of motion information is sufficient to impair the trained forward–backward discrimination ability of the animals, and much more so than for the trained view discriminations.

Both monkeys performed significantly higher on the jerky displays compared to the ISI displays in the forward–backward categorization task (Figure 8A for M1 and Figure 8B for M2). Performance, averaged across the common delay conditions (still frame/ISI durations of 1, 3 and 5, frames) for the jerky locomotions, was 69% and 76% for monkeys M1 and M2, respectively, while it dropped to 55% and 66%, respectively, for the ISI displays (binomial $p < 0.05$, one-sided). Thus, it appears that the introduction of a delay with blanks between subsequent frames has a stronger effect than an increase in spatial displacement. Congruent with the results obtained with the ISI displays, performance on the jerky displays was significantly higher in the view task than in the forward–backward task (for the jerky displays: 96% versus 67% for M1 and 98% versus 78% for M2, both computed for the condition using 3 additional still frames; $p < 0.004$).

The lesser degree of disruption observed in the view task with degraded motion information suggests that perhaps the animals are using only or mainly form information in order to solve this task. However, one should note that the view categorization performance was also affected at long ISIs (although performance was still considerably higher than chance) suggesting that motion

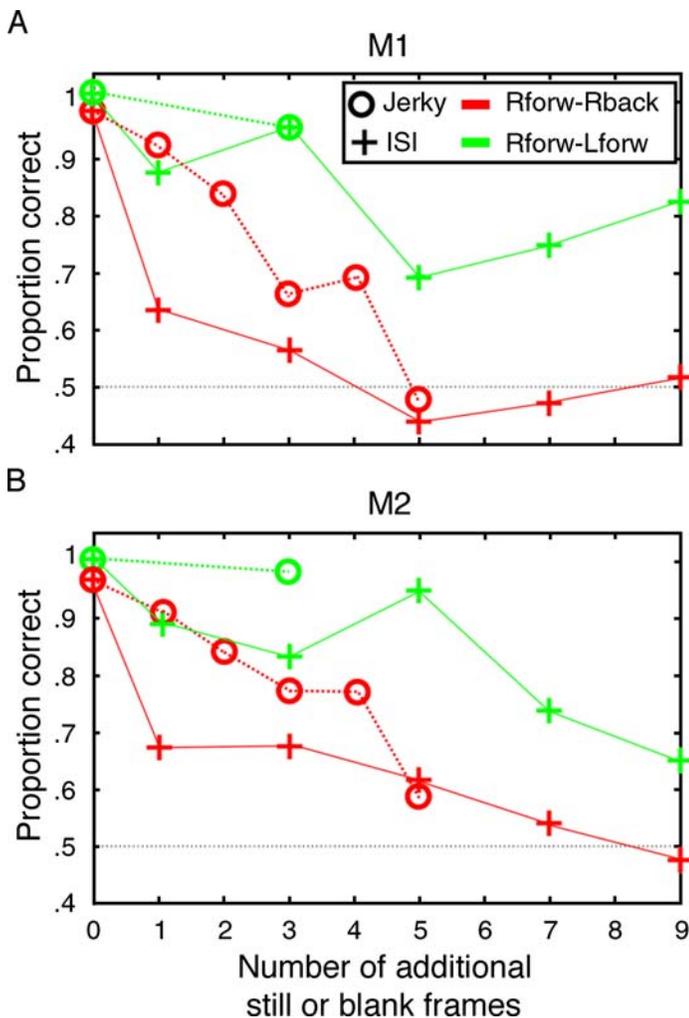


Figure 8. Effect of degrading motion information in the locomotion displays. Proportion correct is plotted as a function of the additional number of still frames for the jerky displays, e.g., locomotion displays with 3 still frames consist of one original snapshot followed by 3 repetitions of the same snapshot (start jerky frame sequence = “1,1,1,1,5,5,5,5,9...”), and as a function of the number of blank frames in the ISI conditions. A value of zero on the X-axis refers to the normal presentation sequence of the standard locomotions. Conventions are the same as in Figure 7; chance levels are 50%. Data are shown for (A) M1 and (B) M2, separately (see also movies: “[movie_fig8_jerkies.mov](#)” and “[movie_fig8_strobos.mov](#)”; note that playing the “jerkies” movie on a CRT display at 60-Hz frame rate is needed to see the disruption of smooth motion).

is also used in the view task. To directly test whether spatial information alone is sufficient to solve the view task, we measured the generalization performance for static presentations of 14 representative body postures. Generalization to static snapshots presented for 300 ms was significant in the 2 monkeys, albeit marginally (binomial test; $p < 0.05$: 59% and 57% correct for M1 and M2, respectively) despite the nearly perfect performance

for the standard locomotion displays (>95% correct). Presenting the static snapshots for 1086 ms yielded similar results. Overall, the weak generalization to the static presentations of the snapshots together with the strong effect of ISI on the view task suggests that these categorizations are not merely based on spatial cues but also depends on motion.

Contribution of opponent motion

Recent psychophysical studies (Casile & Giese, 2005; Thurman & Grossman, 2008) of biological motion using point-light displays suggest that horizontal opponent motion is particularly important for discriminating the direction of locomotion. Therefore, we examined the importance of opponent motion in the discrimination performance of our monkeys by presenting animations consisting of only one leg and one arm, together with the non-opponent body parts such as the head and the torso. Both the view and forward–backward categorizations of these novel displays were assessed in the generalization tests.

Stimuli

One arm–one leg stimuli were rendered for the unshaded humanoids based on the same motion-capture coordinates as those used for generating the standard 4.2 km/h locomotions. All possible combinations of one arm and one leg, together with the non-opponent body parts such as the head and the torso, were created: left arm and left leg, left arm and right leg, right arm and right leg, and right arm and left leg (see movie: “[movie_opponent_motion.mov](#)”).

Tasks

The one arm–one leg stimuli were presented using the same generalization test protocol for the pair of two-alternative tasks described previously. Monkeys M1 and M2 were tested. On average, 28 (M1) and 12 (M2), and 13 (M1) and 8 (M2) generalization trials for the forward–backward and the view task, respectively, were presented for each of the 4 one-arm and one-leg conditions. Note that the standard locomotions were presented in the same test with a frequency of ten times the sum of all generalization trials within that specific test.

Results and discussion

Average generalization performance over the 4 combinations of non-opponent animations was significantly

greater than chance (binomial test; $p < 0.05$): 74% and 85% correct for M1 and M2, respectively, in the forward–backward task and 94% and 84% correct for M1 and M2, respectively, in the view task. Furthermore, significant transfer for each possible combination of legs and arms was observed in both monkeys and both tasks (binomial test; $p < 0.05$ for each combination separately, except for M2 in the view discrimination task with the right arm and right leg displays, most likely a consequence of the small number of repetitions, i.e., 8 repeats). Given these significant transfers to the one arm–one leg displays, we conclude that opponent motion is neither a necessary nor a critical cue for the animals to solve either the view or the forward–backward task. Opponent motion might be a sufficient, but clearly not necessary, cue for the “humanoid” displays that we employed. In addition, the significant generalization to the one arm–one leg locomotions indicate that the presentation of a full, bipedal body is not necessary to obtain significant discrimination of facing direction or forward–backward locomotion.

General discussion

The most striking finding of this behavioral study was the fact that rhesus monkeys needed a much longer training period to distinguish forward from backward walking than for the facing direction/view of the walker. This is remarkable, given the ease with which human observers can categorize both. After extensive training, two monkeys performed very well in the forward–backward task, but performance in this task was highly susceptible to degradation of spatiotemporal coherence and motion information.

Discrimination of forward from backward walking requires the use of motion information, either based on image motion information (Chang & Troje, 2009; Giese & Poggio, 2003: “motion pathway”; Troje & Westhoff, 2006) or based on temporal changes in the body posture (Giese & Poggio, 2003: “form pathway” also designated as “global motion information” by Lange & Lappe, 2006), or both. Categorization of view can be solved using both motion and spatial cues. Thus, the lengthy training needed to distinguish forward from backward locomotions suggests that, unlike humans, monkeys do not spontaneously use such motion information to discriminate locomotion, but instead rely predominantly on spatial differences in trajectories and/or posture-related form cues. It is only after extensive training that the monkeys are able to use the motion information present in the displays.

What sort of motion information did the monkeys in fact use to solve the forward–backward discrimination? The generalization results from the one-legged displays show that opponent motion, as postulated by Casile and

Giese (2005) and Thurman and Grossman (2008), was not necessary, although it may have contributed to the discrimination performance of the “humanoid” displays by the monkeys (note that opponent motion might be important for the human perception of biological motion in more impoverished displays such as point-light stimuli, as suggested by Casile & Giese, 2005; Thurman & Grossman, 2008). As shown in Figure 1B, the direction of motion of the lower limbs differs between the forward and backward trajectories and such local, short-range motion direction signals (Braddick, 1974) might have been used by the animals to solve the task. Perturbation of such local short-range motion signals by modifying the spatiotemporal coherence and/or the ISI can then explain the behavioral effects of these manipulations. The generalization across walking speeds, however, indicates that if such local motion signals indeed contribute, they at least show a speed invariance over a factor of two. A third sort of motion information relates to the “global motion” information postulated for form-based modules of biological motion (Giese & Poggio, 2003; Lange & Lappe, 2006). These modules posit that after matching the form or posture of the walker, using a purely form-driven mechanism, to templates of different postures, forward and backward walking are discriminated by analyzing the sequence in which these postures appear. We cannot exclude this possibility on the basis of our tests. One should note, however, that our data regarding the ISI and jerky-locomotion manipulation differ qualitatively from those of Beintema et al. (2006; see their Experiment 3). The latter authors did not find any differences in performance between jerky locomotions and ISI manipulations when these were matched for stimulus-onset asynchrony (their Figure 5), while our monkeys showed much better performance for the jerky locomotions compared to the ISIs (Figure 8). Perhaps our monkeys were more disturbed by the flicker of the ISI displays than are humans, or perhaps our monkeys simply do not use the form-based mechanism postulated by Beintema et al. (2006) and Lange and Lappe (2006) but rather rely on information from short-range motion mechanisms.

What cues, then, did the animals actually use in our view task? Rforw and Lforw locomotions differ in both motion trajectories and form. Generalization from the trained walking to the running humanoids shows that the learned motion trajectories, if they exist, are relatively broad. The small effects of spatiotemporal coherence, ISI, and jerky-locomotion manipulations on performance in the view task furthermore suggest that motion might not be the predominant cue. Indeed, the animals might have used form cues to distinguish the postural differences between the Lforw and Rforw walkers. As noted before, the view task can be solved, at least in principle, by mere form cues (Beintema et al., 2006; Beintema & Lappe, 2002; Lange & Lappe, 2006). On the other hand, several lines of evidence, i.e., poor generalization to static

snapshot displays and small but significant impairment on the ISI task, suggest that motion information does indeed influence view discrimination performance, at least at high coherences.

The much longer training periods needed to learn the forward–backward compared to the view discriminations might appear to be in conflict with the observation of neurons in the monkey superior temporal sulcus (STS) selective for forward versus backward walking (Oram & Perrett, 1994a, 1994b, 1996). However, in these physiological studies the locomotion stimuli had a translatory component, which was absent in our displays. Like most human psychophysical studies investigating biological motion perception by using tasks containing either view and/or forward versus backward discriminations (e.g., Chang & Troje, 2008, 2009; Hunt & Halper, 2008; Thirkettle et al., 2009; Thurman & Grossman, 2008) we employed stationary locomotions, i.e., treadmill-like walking without a translatory, i.e., trivial, component. One can distinguish intrinsic motion cues, due to changes in body posture during walking, from extrinsic motion corresponding to the translatory motion (e.g., Chang & Troje, 2008). Our displays, using stationary treadmill walkers, contain only intrinsic motion, while the stimuli in the above-mentioned single-cell studies, using real-life presentations in front of passively fixating monkeys, contained both extrinsic and intrinsic motion components. By definition, extrinsic motion trajectories of forward and backward locomotions are very different, particularly for sagittal views of a walker, while the intrinsic motion is more subtle and restricted to the lower limbs. Our results indicate that monkeys do not spontaneously use—at least behaviorally—the intrinsic motion cues related to forward versus backward locomotion but must instead be given extensive training in order to use these cues.

Would the forward–backward categorization have been easier for our monkeys if we had used displays of real human walkers or even monkeys? We do not think that the species factor is important since the present animals are exposed daily to human walkers. In addition, an elegant ethological observation study by Wood et al. (2007) shows that free-ranging rhesus monkeys can recognize human actions (e.g., throwing), even when these actions do not belong to their own behavioral repertoire. Furthermore, kinematic studies in monkeys have shown bipedal locomotion (like our humanoid stimuli) by Japanese monkeys (*Macaca fuscata*; Mori, Nakajima, & Mori, 2006). Bipedalism in monkeys is also present within natural environment settings (Ogihara, Usui, Hirasaki, Hamada, & Nakatsukasa, 2005).

Behavioral animal studies need to be interpreted with care since the particular result may depend on the behavioral task that is employed to assess the perception of the animal. Previous studies have shown that when viewing images of a subject looking into a particular direction, macaques (and human subjects) reflexively orient their attention in the same direction (Deaner &

Platt, 2003; Emery, Lorincz, Perrett, Oram, & Baker, 1997; Lorincz, Baker, & Perrett, 1999). Deaner and Platt (2003) showed that viewing images of a monkey looking, e.g., to the right evoked small rightwards eye position shifts inside a $\pm 2^\circ$ fixation window. One could argue that the animals were suppressing such relatively automatic eye movements (Shepherd, Klein, Deaner, & Platt, 2009) thus increasing the difficulty of the task. However, analyses of eye movements with the degraded walkers showed no significant effects of facing direction during fixation in the first training session. Analysis of aborted trials showed in only two (M2 and M3) of the three animals horizontal eye position differences between aborted Rforw and Lforw conditions consistent with facing direction. Since in our task facing direction and trained saccade direction were not counterbalanced, this difference in eye position in aborted trials is difficult to interpret. Even if such suppression was present in our task, it is extremely unlikely that it can explain the marked difference in training duration of the forward–backward versus view categorization since in both tasks the target point of one of the two discriminants was incongruent with the gaze direction of the walker (LForw required an upward saccade and Rback required a leftward saccade). Note that in the forward–backward task there is a congruency between walking direction and saccade target position, unlike in the view task. One could argue that the former would have even facilitated the learning of stimulus-target associations in the forward–backward task.

One cannot exclude the possibility that our monkeys learned to categorize specific, meaningless patterns of moving cylinders without a link to a representation of a human walker. This could explain both the generalization effects for different actors and the lack of generalization for the differently rendered stimulus formats. The latter is also related to the issue of whether monkeys perceive a meaningful agent in point-light displays of biological motion. So far, attempts to demonstrate recognition by non-human primates of an animate agent in point-light displays have been unsuccessful. After extensively training baboons (*Papio papio*) in discriminating point-light displays of translatory human locomotion from their scrambled counterparts, no transfer to displays of other human actions was found (Parron et al., 2007). In fact, the baboons focused their attention only on subparts of the stimuli. In addition, in a visual search paradigm using point-light displays of locomoting chimpanzees tested in the same species (*Pan troglodytes*) Tomonaga (2001) could not find any search asymmetries specific to the biological motion patterns. However, it should be noted that lack of transfer merely indicates the specificity of the trained categorization, i.e., what is specifically learned in the task, and thus—as any negative result—should be interpreted with caution. In line with the previously mentioned behavioral studies, preliminary results of monkey fMRI experiments performed in our laboratory show that point-light displays of human and/or monkey

locomotion do not activate monkey visual cortex, including the superior temporal sulcus, any more so than do scrambled displays (K. Nelissen, personal communication; J. Jastorff, personal communication). As discussed above, demonstrations of single cell responses in the superior temporal sulcus to point-light walkers used displays that included translation of the walker (Oram & Perrett, 1994a, 1994b). Oram and Perrett (1994b) reported one superior temporal sulcus neuron that responded to a point-light display of a non-rigidly rotating body without translatory motion, but it is not clear which motion features this neuron responded to.

The present study demonstrates a difference between monkey and human subjects in the perception of locomotion displays more abstract than those of real humans or other animals. Such reduced stimuli are widely used in human research on biological motion and links are often made to neurophysiological data obtained in non-human primates. However, natural and effortless discrimination of locomotion displays in humans—such as forward from backward walking—does not necessarily imply that monkeys also make such discriminations easily, as shown in the present study. Several other recent studies have documented differences in the processing of complex visual stimuli between humans and monkeys (e.g., Einhäuser, Kruse, Hoffman, & Köning, 2006; Nielsen, Logothetis, & Rainer, 2008) implying that, especially for complex stimuli, any extrapolation of human findings to monkey perception, and vice versa, always needs to be made with caution.

Although it required extensive training, eventually the animals were able to discriminate forward from backward locomotions almost perfectly. Thus the question arises as to which region and which neural selectivity underlies the successful forward–backward categorization obtained through training. Oram and Perrett (1994a, 1994b, 1996) demonstrated neural selectivities for forward versus backward walking in the rostral superior temporal sulcus of macaques. However, as mentioned above, the main real-life displays used in these single cell studies contained a strong translatory, i.e., extrinsic, motion component and thus it remains to be investigated whether similar neural selectivities exist using the stationary and well-controlled locomotions of this study, which contain only intrinsic motion. Vangeneugden et al. (2009) distinguished between two types of neural selectivities to simple arm action displays in the rostral monkey superior temporal sulcus: neurons that respond as well to static snapshot displays of the action and to the action itself, dubbed “snapshot” neurons predominantly located in the ventral bank of the STS, and neurons that respond more strongly to actions than to the static snapshot displays, dubbed “motion” neurons predominantly located in the dorsal bank of the STS. It is tempting to speculate that both “snapshot” and “motion” neurons contribute to the categorization of locomotion direction, since these displays differ in both form and motion cues, while it is

mainly “motion” neurons that contribute to forward–backward categorizations.

Acknowledgments

The technical assistances of P. Kayenbergh, G. Meulemans, M. De Paep, W. Depuydt, S. Verstraeten, I. Puttemans, and M. Docx are gratefully acknowledged. Péter Kaposvári assisted in the training of the animals. We thank S. Raiguel for reading a previous version of the manuscript and also two reviewers for their helpful comments. This research was supported by Geneeskundige Stichting Koningin Elizabeth, GOA/2005/18, Detection and Identification of Rare Audiovisual Cues (DIRAC) FP6-IST 027787, EF/05/014, and IUAP. J.V. was a research assistant of the Fund for Scientific Research Flanders (FWO: Fonds voor Wetenschappelijk Onderzoek Vlaanderen).

Commercial relationships: none.

Corresponding author: Rufin Vogels.

Email: Rufin.Vogels@med.kuleuven.be.

Address: Laboratorium voor Neuro-en Psychofysiologie, K.U. Leuven Medical School, Herestraat 49, B-3000 Leuven, Belgium.

References

- Beintema, J. A., Georg, K., & Lappe, M. (2006). Perception of biological motion from limited-lifetime stimuli. *Perception & Psychophysics*, *68*, 613–624. [[PubMed](#)] [[Article](#)]
- Beintema, J. A., & Lappe, M. (2002). Perception of biological motion without local image motion. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 5661–5663. [[PubMed](#)] [[Article](#)]
- Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, *58*, 47–73. [[PubMed](#)]
- Braddick, O. (1974). A short-range process in apparent motion. *Vision Research*, *14*, 519–527.
- Casile, A., & Giese, M. A. (2005). Critical features for the recognition of biological motion. *Journal of Vision*, *5*(4):6, 348–360, <http://journalofvision.org/5/4/6/>, doi:10.1167/5.4.6. [[PubMed](#)] [[Article](#)]
- Chang, D. H. F., & Troje, N. F. (2008). Perception of animacy and direction from local biological motion signals. *Journal of Vision*, *8*(5):3, 1–10, <http://journalofvision.org/8/5/3/>, doi:10.1167/8.5.3. [[PubMed](#)] [[Article](#)]
- Chang, D. H. F., & Troje, N. F. (2009). Characterizing global and local mechanisms in biological motion perception. *Journal of Vision*, *9*(5):8, 1–10, <http://journalofvision.org/9/5/8/>

- journalofvision.org/9/5/8, doi:10.1167/9.5.8. [PubMed] [Article]
- Churchland, M. M., & Lisberger, S. G. (2001). Shifts in the population response in the middle temporal visual area parallel perceptual and motor illusions produced by apparent motion. *Journal of Neuroscience*, *21*, 9387–9402. [PubMed] [Article]
- Clarke, F. R. (1957). Constant ratio rule for confusion matrices in speech communication. *Journal of the Acoustical Society of America*, *29*, 715–720.
- Deaner, R. O., & Platt, M. L. (2003). Reflexive social attention in monkeys and humans. *Current Biology*, *13*, 1609–1613. [PubMed] [Article]
- Einhäuser, W., Kruse, W., Hoffman, K. P., & Köning, P. (2006). Differences of monkey and human overt attention under natural conditions. *Vision Research*, *46*, 1194–1209. [PubMed]
- Emery, N. J., Lorincz, E. N., Perrett, D. I., Oram, M. W., & Baker, C. I. (1997). Gaze following and joint attention in rhesus monkeys (*Macaca mulatta*). *Journal of Comparative Psychology*, *111*, 286–293. [PubMed]
- Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews, Neuroscience*, *4*, 179–192. [PubMed]
- Grossman, E. D. (2005). Evidence for a network of brain areas involved in perception of biological motion. In M. Grosjean, G. Knoblich, M. Shiffrar, & I. M. Thornton (Eds.), *The human body: Perception from the inside out* (pp. 361–384). Oxford, UK: Oxford University Press.
- Hunt, A. R., & Halper, F. (2008). Disorganizing biological motion. *Journal of Vision*, *8*(9):12, 1–5, http://journalofvision.org/8/9/12, doi:10.1167/8.9.12. [PubMed] [Article]
- Jhuang, H., Serre, T., Wolf, L., & Poggio, T. (2007). A biologically inspired system for action recognition. *Proceedings of the Eleventh IEEE International Conference on Computer Vision (ICCV)*.
- Lange, J., & Lappe, M. (2006). A model of biological motion perception from configural form cues. *Journal of Neuroscience*, *26*, 2894–2906. [PubMed] [Article]
- Lorincz, E. N., Baker, C. I., & Perrett, D. I. (1999). Visual cues for attention following rhesus monkeys. *Current Psychology of Cognition*, *18*, 973–1003.
- MacMillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. New Jersey: Lawrence Erlbaum.
- Mather, G., Radford, K., & West, S. (1992). Low-level visual processing of biological motion. *Proceedings of the Royal Society of London B: Biological Sciences*, *249*, 149–155. [PubMed] [Article]
- Mikami, A., Newsome, W. T., & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex: I. Mechanisms of direction and speed selectivity in extrastriate area MT. *Journal of Neurophysiology*, *55*, 1308–1327. [PubMed]
- Mori, F., Nakajima, K., & Mori, S. (2006). Control of bipedal walking in the Japanese monkey, *M. fuscata*: Reactive and anticipatory control mechanisms. In H. Kimura, K. Tsuchiya, A. Ishiguro, & H. Witt (Eds.), *Adaptive motion of animals and machines* (pp. 249–259). Tokyo, Japan: Springer Tokyo.
- Newsome, W. T., Mikami, A., & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex: III. Psychophysics and physiology of apparent motion. *Journal of Neurophysiology*, *55*, 1340–1351. [PubMed]
- Nielsen, K. J., Logothetis, N. K., & Rainer, G. (2008). Object features used by humans and monkeys to identify rotated shapes. *Journal of Vision*, *8*(2):9, 1–15, http://journalofvision.org/8/2/9, doi:10.1167/8.2.9. [PubMed] [Article]
- Ogihara, N., Usui, H., Hirasaki, E., Hamada, Y., & Nakatsukasa, M. (2005). Kinematic analysis of bipedal locomotion of a Japanese macaque that lost its forearms due to congenital locomotion. *Primates*, *46*, 11–19. [PubMed]
- Oram, M. W., & Perrett, D. I. (1994a). Neural processing of biological motion in the macaque temporal cortex. *Proceedings of SPIE*, *2054*, 155–165.
- Oram, M. W., & Perrett, D. I. (1994b). Responses of anterior superior temporal polysensory (STPa) neurons to “biological motion” stimuli. *Journal of Cognitive Neuroscience*, *6*, 99–116.
- Oram, M. W., & Perrett, D. I. (1996). Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *Journal of Neurophysiology*, *76*, 109–129. [PubMed]
- Parron, C., Deruelle, C., & Fagot, J. (2007). Processing of biological motion point-light displays by baboons (*Papio papio*). *Journal of Experimental Psychology: Animal Behavior Processes*, *33*, 381–391. [PubMed]
- Schindler, K., & Van Gool, L. (2008). Action snippets: How many frames does human action recognition require? *Proceedings of the IEEE Conference in Computer Vision and Pattern Recognition (CVPR)*.
- Shepherd, S. V., Klein, J. T., Deaner, R. O., & Platt, M. L. (2009). Mirroring of attention by neurons in macaque parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 9489–9494. [PubMed] [Article]
- Thirkettle, M., Benton, C. P., & Scott-Samuel, N. E. (2009). Contributions of form, motion and task to biological motion perception. *Journal of Vision*,

- 9(3):28, 1–11, <http://journalofvision.org/9/3/28>, doi:10.1167/9.3.28. [[PubMed](#)] [[Article](#)]
- Thornton, I., Pinto, J., & Shiffrar, M. (1998). The visual perception of human locomotion. *Cognitive Neuropsychology*, *15*, 535–552.
- Thurman, S. M., & Grossman, E. D. (2008). Temporal “bubbles” reveal key features for point-light biological motion perception. *Journal of Vision*, *8*(3):28, 1–11, <http://journalofvision.org/8/3/28>, doi:10.1167/8.3.28. [[PubMed](#)] [[Article](#)]
- Tomonaga, M. (2001). Visual search for biological motion patterns in chimpanzees (*Pan troglodytes*). *Psychologia*, *44*, 46–59.
- Troje, N. F., & Westhoff, C. (2006). The inversion effect in biological motion perception: Evidence for a “life detector?”. *Current Biology*, *16*, 821–824. [[PubMed](#)] [[Article](#)]
- Vangeneugden, J., Pollick, F., & Vogels, R. (2009). Functional differentiation of macaque visual temporal cortical neurons using a parametric action space. *Cerebral Cortex*, *19*, 593–611. [[PubMed](#)]
- Vogels, R. (1999). Categorization of complex visual images by rhesus monkeys: Part 1. Behavioral study. 1223–1238. [[PubMed](#)]
- Wood, J. N., Glynn, D. D., & Hauser, M. D. (2007). The uniquely human capacity to throw evolved from a non-throwing primate: An evolutionary dissociation between action and perception. *Biology Letters*, *3*, 360–364. [[PubMed](#)] [[Article](#)]