Depth perception from point-light biological motion displays

Marc H. E. de Lussanet
Institute for Psychology, University of Münster, Münster, Germany

Markus Lappe
Institute for Psychology, University of Münster, Münster, Germany

Humans have a clear impression of facing in depth for point-light biological motion. However, this has not been measured systematically nor is it known on which cues humans rely for their judgment. In the present study subjects judged the facing orientation-in-depth of point-light displays. The displays represented natural walking and modified versions in which the time sequence was reversed, action was perturbed, the limbs and joints were nonrigid, the temporal sequence was scrambled, or the joint positions were scrambled. We found that the subjects were best at judging the facing direction of normal and reversed walking with an accuracy of 6° and 10° precision. The results show that pendular motion of the limb segments and the implicit knowledge of the human body play an important role for the precision of the judgment. Three further factors were relevant for the judgment of facing direction: (a) the discrimination of the front and back side, (b) the facing bias, and (c) the impression of depth from the display, probably due to the kinetic depth effect. The latter influences the accuracy, which differed strongly between subjects. The results suggest that the facing bias, to perceive the figure as facing toward the observer rather than away, is not related to the recognition of a human figure but rather to the presence of oscillating movements of the dots in the display.

Introduction

Point-light biological motion is possibly the most abstracted and impoverished visual display of actions, and humans are remarkably good at its recognition (Johansson, 1973). Typical studies have used profile views of walking, but human observers have a strong impression of depth if an oblique or frontal view is displayed (Bülthoff, Bülthoff, & Sinha, 1998). In normal life, walking may indeed be one of the most common whole-body activities that we may see, but a pure side view is rather uncommon. Instead, it is relevant to perceive in which direction a person is heading or which orientation the person has to the observer.

Perception of 3D orientation has been systematically studied for objects and faces. For example, rigid 3D objects and faces can be best recognized from canonical views (Blanz, Tarr, & Bülthoff, 1999), and humans tend to choose preferred views when learning 3D objects and faces (Harries, Perrett, & Lavender, 1991; Peissig & Tarr, 2006; Perrett & Harries, 1988). It has also been shown that local and global strategies are combined in a flexible manner when judging the facing in depth of 3D objects (Foster & Gilson, 2002; Troje & Bülthoff, 1996; Watson, Johnston, Hill, & Troje, 2005). Different views of objects and faces are represented by different populations of cells in the ventral temporal cortex in humans (Grill-Spector et al., 1999). In monkeys it has been found that these populations are typically organized in orientation maps (Wang, Tanifuji, & Tanaka, 1998).

For point-light biological motion, it has been found that some neurons in the monkey inferotemporal cortex are selective for the 3D orientation (facing in depth) (Vangeneugden et al., 2011). Also, in humans there is evidence for such orientation maps (Michels, Kleiser, de Lussanet, Seitz, & Lappe, 2009). Psychophysical experiments showed that humans can detect changes in the orientation in depth even if they occur during a saccade (Verfaillie & de Graef, 2000). Humans can also easily discriminate between forwards and reversed walking from almost every viewpoint even from single frame lifetime point lights (Kuhlmann, de Lussanet, & Lappe, 2009). An exception to this is the frontal view in which the performance was quite bad and depended entirely on the presence of the foot-point lights in the display.

Two kinds of depth information are available for seeing depth from a point light display, which we shall address as implicit and explicit depth cues. With *explicit depth cues* we mean cues that can be present in any kind of visual stimulus, such as occlusion,

binocular disparity, and perspective deformations. Explicit depth cues do contribute to the perception of depth (Vanrie, Dekeyser, & Verfaillie, 2004). With *implicit depth cues* we mean cues for which implicit knowledge of the human body and its movements is needed. Current theories of biological motion perception are based on such implicit knowledge (Giese, 2004; Lange & Lappe, 2006; Lee & Wong, 2004). An interesting feature of implicit depth cues in point-light displays is that they are ambiguous for symmetric actions. This means that a frontal view of walking is the same as a back view (see also the more detailed explanation in the Methods section, Figure 2C). In the present study, we will only address implicit depth cues.

Humans have a clear impression of whether a point-light display faces toward or away from the observer, even if explicit depth cues are absent (Brooks et al., 2008; Jackson, Cummins, & Brady, 2008; Schouten, Troje, Brooks, van der Zwan, & Verfaillie, 2010; Schouten, Troje, & Verfaillie, 2011; Schouten, Troje, Vroomen, & Verfaillie, 2011; Sweeny, Haroz, & Whitney, 2012; Vanrie et al., 2004; Vanrie & Verfaillie, 2006). Explicit depth cues can disambiguate the point-light display and thus determine which of the two ambiguous facing directions an observer sees (Jackson & Blake, 2010; Vanrie et al., 2004).

However, it is unknown how well humans can actually judge the facing in depth of point-light biological motion, and neither is it known on the basis of what sort of implicit information they do so. The answers to these questions are relevant because they may help understand the underlying mechanisms that the brain uses to recognize biological motion. Such information may include dynamic and static cues. A number of dynamic cues have been proposed for recognizing a side view of point-light walking, which could in principle be of help to recognize the facing in depth.

First, we have the knowledge that human movements are typically pendular, in the sense that the limb segments are rigid so that the movement patterns of the major joints typically revolve with respect to the adjacent joints (Chang & Troje, 2009; Hoffman & Flinchbaugh, 1982; Jokisch & Troje, 2003; Webb & Aggarwal, 1982). Second, we have implicit knowledge of the trajectories of individual body parts such as wrists and ankles (Troje & Westhoff, 2006). Third, we have implicit knowledge of the entire movement pattern (Giese, 2004). Fourth, it has been proposed that humans use their implicit knowledge of the body form at each stage of a movement, which is a motion-independent cue (Beintema & Lappe, 2002; Lange, Georg, & Lappe, 2006). Finally, local dynamic cues (Neri, 2009), or local static cues such as the ratio of the distance between the hips or shoulders over the length of limb segments, could be informative.

In the present study, we aimed to measure how well human observers can tell the facing in depth in the absence of explicit depth cues and to extract the relative importance of the different implicit cues mentioned above. Five manipulations were applied to a recorded pattern of human walking. These manipulated systematically the naturalness of the motion, the rigid structure of the body segments, the body structure, the movement of the dots, and the underlying coherency of the points. These stimuli faced in a random horizontal direction. The task of the subjects was to report the facing direction as precisely as possible. In the Discussion we will address the question of which mechanisms might underlie the depth perception from implicit depth cues on the basis of the kinds of errors that the observers make.

Methods

Subjects

The nine participants were undergraduate students from the psychology department. They received credit points for experimental participation in the regular undergraduate program. All participants had normal or corrected to normal vision and no known history of neurological disorders.

Stimuli and setup

The stimuli were computed from a walking cycle of a 20-year old female with a regular, relatively symmetric gait (note that most people's gait is surprisingly asymmetric). A MotionStar (Ascension) motion capture system recorded the 3D motions of markers attached to the body. The horizontal translation was subtracted. Then a single, regular gait cycle (beginning at right heel ground contact) was selected. The small mismatches of the start and end of each joint's trajectory were removed by adding linear gains. The recordings were then smoothed to remove high-frequency recording noise.

The bounding box (the 3D volume that just contained the trajectories of all points) was computed as well as the length of the upper and lower limb segments and the hip and shoulder segments. The hip- and shoulder-joint-angles were defined between the upper limb segment and the trunk. The knee, elbow, hip, and shoulder joint angles as well as the orientation of the trunk, hip, and shoulder segments over time were computed. See also the Supplementary Quicktime movies.

The normal condition was an orthogonally projected view of the point-light walking stimulus without
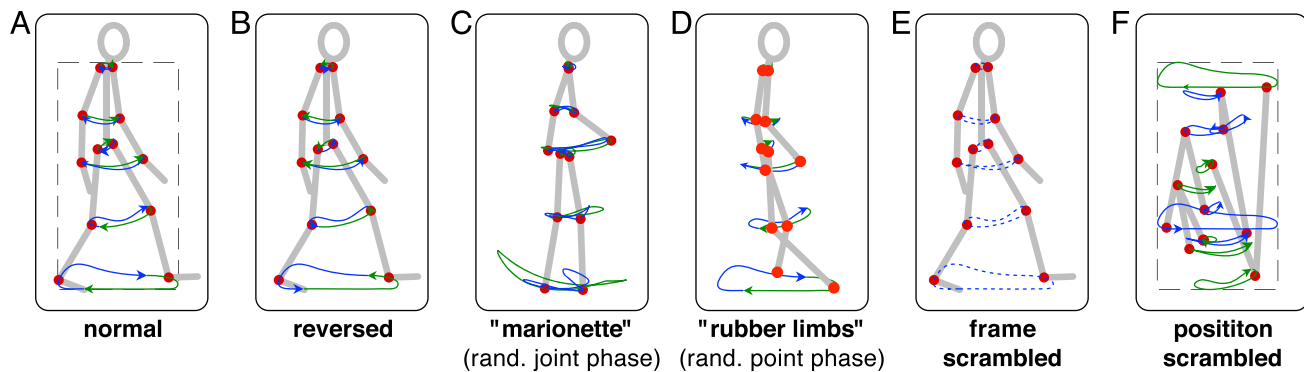
Figure 1. Schemas of the different point-light stimuli in side-view (0°). In grey the underlying body structure. Blue and green depict the trajectories of the visible point-lights. The real stimuli consisted only of the point-lights. (A) Normal walking. (B) Reversed walking (cf. arrow heads in the trajectories). (C) In marionette, the body structure and joint angles are natural but each joint of a pair follows a different, unusual trajectory due to a phase offset. (D) In rubber limbs, the joints follow the normal trajectories but at the wrong time. The body segments are nonrigid and the joints bend in unnatural directions. (E) In frame-scrambled, the human figure is intact. There is no apparent motion and the point-lights do not follow trajectories because a random phase is presented at every frame. (F) In position-scrambled, the joints follow their intact trajectories at the correct time but the wrong location. The body structure and relative movements are destroyed.

occlusion. (Figure 1A). The following scrambled versions were all computed from the normal stimulus.

Reversed was the normal cycle played backwards (Figure 1B).

For marionette, the structure of the body was left intact: The limb segments all had the same lengths and the joints bended through the same range as during normal walking (Figure 1C). The manipulation shifted the timing of the bending of the joints. The stimulus was therefore biomechanically possible at all times, whereas the movement kinematics were very unnatural. Also, the joints of the left and right side of the body followed different trajectories. The stimulus thus gave the impression of a marionette.

For rubber limbs, the spatial trajectories of the joints were left intact but the timing of these trajectories was shuffled (Figure 1D). Consequently, the limb segments changed their lengths during the walking cycle and the knees and elbows could adopt angles in the non-physiological range. The resulting stimulus was biomechanically impossible because the limb segments were not rigid and because the joints bent in non-physiological directions. This gave the stimulus a nonhuman, rubber-like impression.

In frame-scrambled, on each time frame a random phase of the normal stimulus was displayed so there were no trajectories and no apparent motion of the points (Figure 1E).

For the position-scrambled stimulus each point followed the normal trajectory and phase but with a random offset in 3D (Figure 1F). These offsets were such that the bounding box was the same size as the normal stimulus. This stimulus does not look human and uninformed observers would never guess that the stimulus is related to human movement in any way.

The relationships between the joints of one limb are not recognizable.

Table 1 summarizes how the stimuli are gradually more different from the normal stimulus. Listed are the implicit, stimulus-related cues that are available for estimating the facing in depth.

Procedure

The participant was acquainted with the six different stimulus types. It was then explained that he or she was to judge the exact direction in depth in which the stimulus was facing. After two full cycles (2.76 s), the stimulus was replaced by a circular, mouse-controlled cursor (Figure 2A). The cursor could be rotated towards the desired direction by dragging the mouse. When the cursor direction corresponded with the perceived facing in depth the participant pressed the space bar and the next stimulus appeared.

Each of the six stimulus types (Figure 1) was presented 60 times. Each stimulus was simulated with a random facing in depth (α). The trials were blocked by stimulus type. The order of the stimulus types was balanced between the participants.

Analysis

Three factors are essential to analyze the results: the facing bias, the accuracy, and the precision.

The facing bias

First, the facing-towards-the-observer bias was computed as the percentage of responses between 0°

Stimulus	Human form	Human movement	Pendular motion	Point trajectories	Local motions
Normal	Correct	Correct	Correct	Correct	Correct
Reversed	Correct	Dynamically incorrect	Correct	Correct	Dynamically incorrect
Marionette	Correct	Dynamically incorrect	Correct	Dynamically incorrect	Dynamically incorrect
Rubber	Physically impossible	Physically impossible	Physically impossible	Correct	Correct
Frame-scrambled	Correct	–	–	–	–
Position-scrambled	–	–	Physically impossible	Correct	Correct

Table 1. Presence of implicit cues that might be used for judging the facing in depth in the stimuli used in the experiment. Dynamically incorrect: The stimulus is physically consistent with the human body but not normally performed this way. Physically incorrect: The cue is Physically impossible, e.g., the joints bend in unnatural directions and limb segments are non-rigid. A – symbol: The cue is absent.

and 180°. To avoid a possible influence from the stimuli of the previous block, only the last 50 trials of each block were used for this measure.

A profile view neither faces into the screen nor out of it, so it has no bias. Therefore, only trials in which both the facing direction of the stimulus and the response deviated more than 1.0° from a side view were included (84% for frame-scrambled, 90% for position-scrambled, and 94% for the other conditions).

The accuracy or mean error

The mean error was computed as a measure of the accuracy. For this, a positive error was defined in the direction away from the plane of the display screen (grey areas in Figure 2D). Thus, for a side view (i.e., 0° or 180°), any errors were positive by definition and in the range 0°–90°. For a front-back view (i.e., 90° or 270°), any errors were negative by definition, and in the range 0° to −90° (white range in Figure 2D). For example, if a stimulus at 45° was responded with 48°, this would be an error of +3°. If the same stimulus was perceived as facing away (i.e., 315°), an error of +3° would be obtained by responding 312°. Note that this definition has the advantage that it is symmetric with respect to the midline, so that errors in the two hemifields are directly comparable and can be averaged.
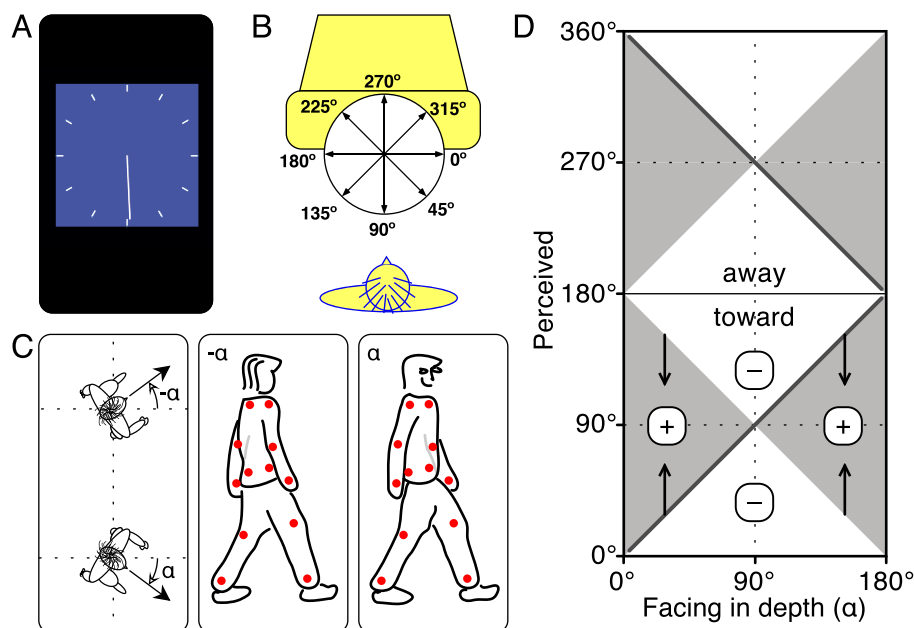

Figure 2. Definition of facing angle α. (A). Screen shot of the response mask. The white arrow (currently pointing at 88° out of the plane of screen) was controlled by the mouse. (B) Schematic top view of the experimental setting, showing the correspondence of the setting in the response mask and the perceived depth orientation. (C) Any given configuration of points corresponds with two depth orientations, −α and α, which are mirrored in the plane of the display screen. (D) The relationship between the angle of facing in depth, perceived angles, and errors. Correct responses are located on the diagonal grey lines. Since α represents the same configuration as −α, two responses are correct for any facing angle (except the side views). Errors towards the midplane (90° and 270°) are positive (grayed regions; see text).
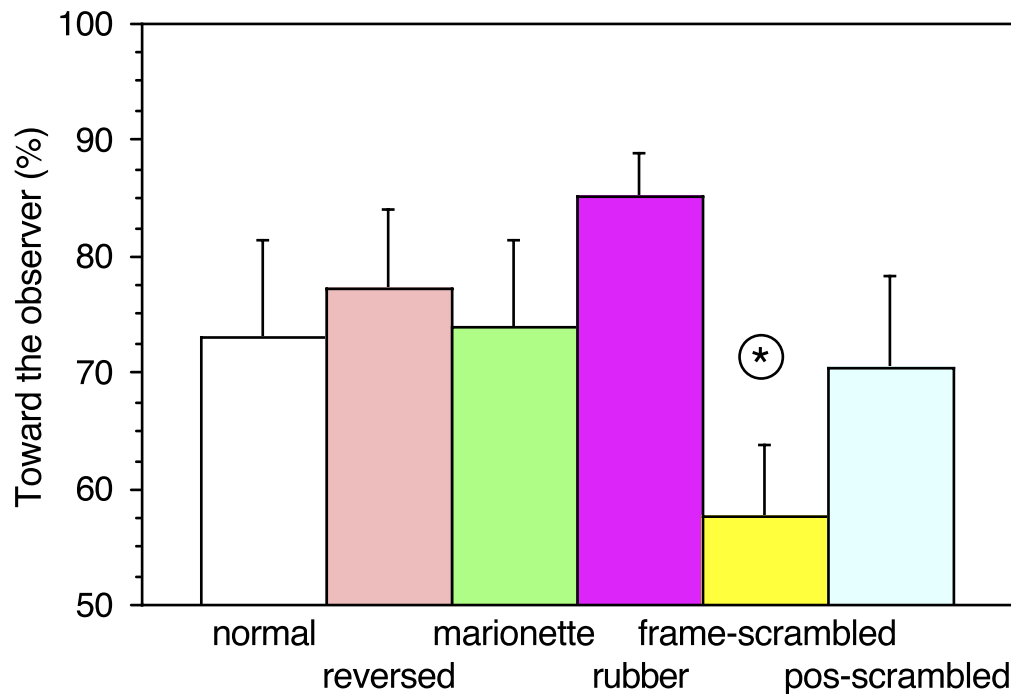
Figure 3. Facing-the-observer-bias. The 100% is a maximal bias; at 50% there is no bias. The asterisk indicates a significant difference from all other conditions.

Also note that the measure is insensitive to a confusion of the front and backside of a stimulus. This can occur, for example, with the position-scrambled stimulus (Figure 1F) which does not have a uniquely defined front or backside. If we apply the example, while the observed confuses front and back, he would not respond at 48°, but at 180° + 48° = 228°, which would still be an error of 3° (with respect to 225°, the left-oblique away direction).

The precision or standard deviation

As a measure of the precision of the settings, the standard deviation of the error was computed. As a reference for the reader: A homogeneous random distribution of responses would result in a standard deviation of 37.0°, whereas the expectancy for giving the same response on each trial (e.g., the plane of the display) is 26.2°.

Statistics

The data were plotted with the package R (Version 2.13, with the R.app GUI for Mac). A locally weighted scatterplot smoothing (LOWESS) moving-average with a window of 10° was fitted to the data of all subjects for each stimulus type.

For statistical comparisons between conditions, repeated measures analyses of variance (ANOVAs) were computed (with stimulus type as a within-factor). For the planned pairwise comparisons, Fisher's partial least squares difference test (PLSD) was used. In a number of cases, the results were also tested against chance level or the expectancy level for random responses. In these cases signed t tests were used with a Bonferroni-Holm correction for multiple comparisons (Holm, 1979).

Results

Facing-the-observer bias

The percentage of facing-the-observer responses is presented in Figure 3. An ANOVA with scramble type as within-factor revealed a significant effect of scramble type, $F(5, 48) = 4.5$, $p = 0.003$. Fisher's PLSD tests confirmed that frame-scrambled differed significantly from all other conditions ($p < 0.02$, Figure 3).

T tests showed that the frame-scrambled condition was the only one that did not differ significantly from the 50% chance level ($p = 0.3$). All the other conditions were significantly above 50% ($p < 0.02$; Bonferroni-Holm correction). Thus, in all conditions but the frame-scrambled condition, the observers had a systematic bias to see the stimuli as facing out of the plane of the screen.

Error distributions

Figure 4 plots the responses for each of the stimuli as a function of the simulated facing in depth. Correct
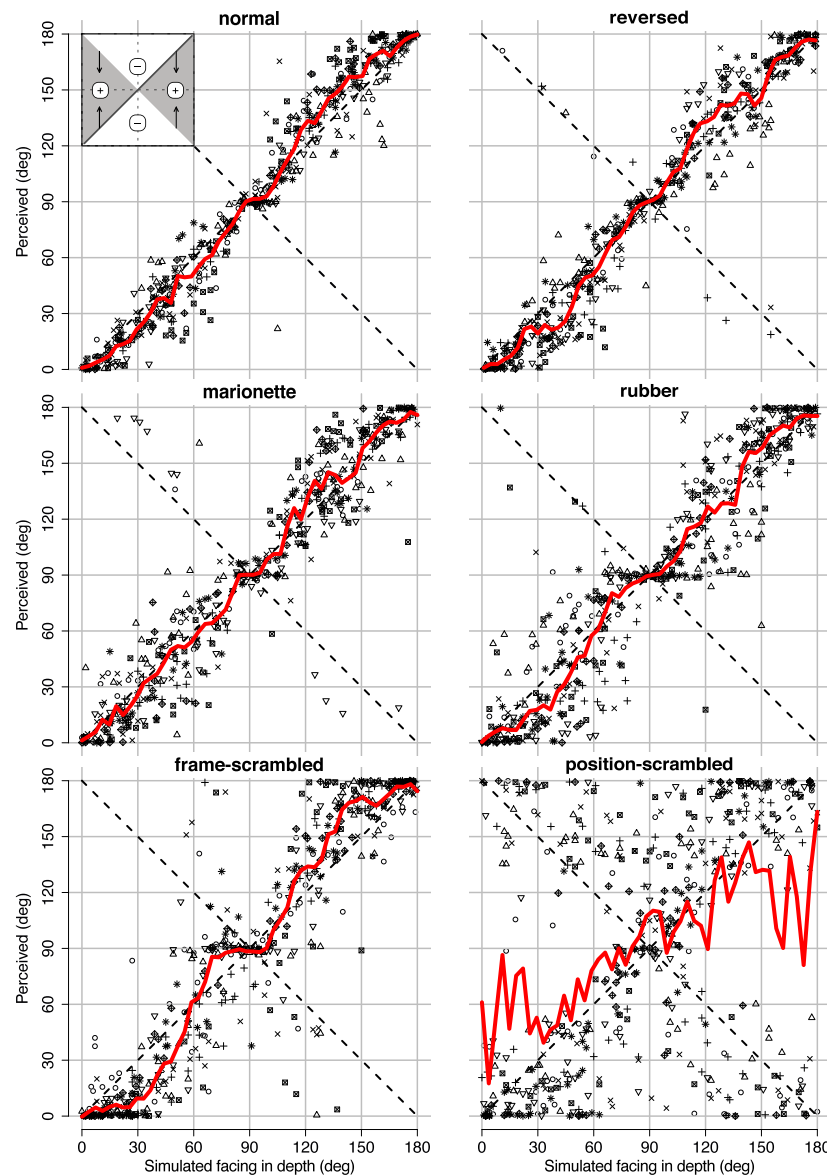
Figure 4. Distribution of the responses in each condition as a function of the simulated facing in depth (α). Each subject is represented by a different symbol. Note that responses between 180°–360° are converted to 180°–0° (cf. Figure 2D and the inset in the upper left panel). 0°: facing right. Red curves: LOWESS moving-average fit with a window of 10°. Ascending diagonal: correct responses; descending diagonal: wrong facing direction.

responses are located on the ascending diagonal, whereas the opposite facing direction is depicted by the descending diagonal. The latter kind of errors can occur if the front and backside are confused. Such errors were rare and most common in the reversed and marionette conditions.

A different kind of error is the tendency to respond along the cardinal axes (0°, 90°, 180°, and 270°). These responses are clustered around the horizontal 0°, 90°, and 180° lines. These errors were most common for the rubber and frame-scrambled conditions, with a probability of a response within 3° of a cardinal axis of about five times the expectancy rate of a random

distribution. For the normal, reversed, and marionette stimuli this probability was about 2.5 times the expected rate.

In most conditions the moving average curves show a slight S-shape: In the range of 0°–90° the curves are slightly below the diagonal whereas for 90°–180° the curves are somewhat higher. This indicates that the observers tended to underestimate how much the facing in depth deviated from the plane of the display. Due to the S-shape, the errors tend to be negative (cf. Figure 2D). In other words, the distributions indicate that the precision tends to be negative.
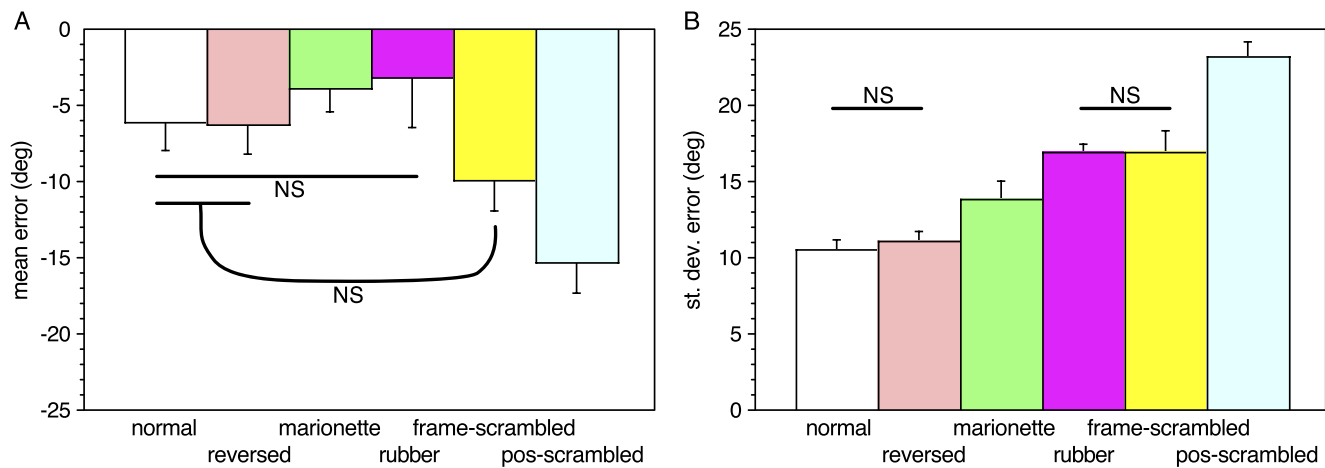
Figure 5. (A) Mean error (accuracy) and (B) the standard deviation (precision) of the error. NS: the only nonsignificant differences (see text). Error bars: standard error.

Accuracy and precision (mean and variable errors)

Figure 5 presents the two error measures. Note that the two panels have the same scale. The error bars are much larger for the mean errors (the accuracy, Panel A) than for the variable errors (the precision, Panel B). This reflects that the S-shape in the distributions of Figure 4 (see above) was strongly expressed for some subjects, but not for all.

The S-shape in the response distributions resulted in a negative mean error of $-6°$ for the normal stimuli (the accuracy, Figure 5A). The ANOVA on the mean errors showed a significant main effect of scrambling type, $F(5, 48) = 10.6$, $p < 0.0001$. Fisher's PLSD post-hoc tests revealed that the differences between position-scrambled and all other conditions, as well as between frame-scrambled and the rubber and marionette conditions, were statistically significant ($p < 0.01$). The normal condition differed only significantly from the position-scrambled condition. T tests showed that the mean errors differed significantly from zero in the normal reversed, frame-scrambled, and position-scrambled conditions ($p < .01$, corrected) and did not in the rubber condition ($p < 0.4$).

The ANOVA on the variable errors (the precision, Figure 5B) also showed a significant main effect of scrambling type, $F(5, 48) = 44.0$, $p < 0.0001$. Fisher's PLSD post-hoc tests showed that the only conditions that did not differ significantly were normal and reversed and rubber and frame-scrambled ($p > 0.5$). All other differences were highly significant ($p < 0.01$). All conditions differed significantly from zero ($p < 0.01$, corrected t tests). As explained in the Methods, the optimal guessing strategy, i.e., giving the same response on every trial, would result in a precision of $26.2°$. The precision was significantly lower than $26.2°$

in all conditions ($p < 0.05$, corrected) except the position-scrambled stimuli ($p = 0.06$).

The mean and variable errors of subjects were not correlated for any of the six conditions ($r < 0.26$, $p > 0.5$).

Control experiment: Frame duration is irrelevant for frame scrambling

The frame-scrambled condition of main experiment was presented with a 10-ms frame duration, which very short and may have deteriorated the performance of participants. The control experiment included only the frame-scrambled conditions but with a longer frame duration of 80 or 160 ms. With longer frame durations the subsequent configurations might have been recognized better and thus improved performance.

A new group of 11 undergraduate students performed in the control experiment. Each still configuration of the frame-scrambled stimulus remained on the screen for either eight or 16 monitor refreshes (i.e., 80 or 160 ms).

The repeated-measures ANOVAs on the mean errors, on the standard deviation of the error, and on the percentage of toward-responses did not reveal any significant main effects of frame duration, $F(1, 20) < 3$, $p > 0.1$. These measures were all very similar to those found for the scrambled frame-order condition in the main experiment (mean error $= -8.9°$, standard deviation error $= 17.0°$, towards responses $= 47\%$). Three new ANOVAs with experiment as a between-factor confirmed this: There was no significant main effect of experiment for any of the three measures, $F(1, 28) < 1.0$, $p > 0.3$.

The results of the control experiment confirm that the frame duration was not a limiting factor for the

recognition of the facing direction in the main experiment. The results also reproduce the lack of a facing bias in the frame-scrambled condition.

Discussion

The results of the present study show that non-trained human observers can accurately judge the facing in depth from two-dimensional point-light biological motion displays. As we will discuss below, both, the familiar structure of the body and the relative *pendular motions* of the joints contributed to this.

The results revealed four different factors to the judgment of the facing in depth of a point-light display. These are (a) the discrimination of the front and back side, (b) the facing bias, (c) the impression of depth from the display, and (d) the precision of the judged facing direction.

Confusion of front and back

The participants were generally very good at discriminating the front and back sidse (cf. Figure 4, in all conditions but the position-scrambled there were far more responses around the ascending dashed lines than around the descending dashed lines). This is in line with earlier findings in which the task was only to discriminate between facing left and right (Beintema & Lappe, 2002, for normal walking; Lange & Lappe, 2007, for frame-scrambled). The position-scrambled condition showed a substantial number of confusions of the front side and back side, consistent with the finding of Troje and Westhoff (2006). The present study extends these findings to facing in depth orientations.

Facing bias

We found a strong facing-toward-the-observer bias in most conditions. This finding corroborates earlier findings (Brooks et al., 2008; Manera et al., 2012; Vanrie et al., 2004; Vanrie & Verfaillie, 2006). These studies all investigated the depth ambiguity inherent to biological motion without explicit depth cues. In some cases, such as walking, this ambiguity leads to a strong facing bias (Vanrie et al., 2004; Vanrie & Verfaillie, 2006). More generally, this kind of bias is known as depth inversion (Yellott & Kaiwi, 1979). A well known example from research on face perception is the hollow face illusion (Hill & Johnston, 2007).

Many biological motion studies on the facing bias have sought the reason for this bias in social or action related explanations. Vanrie and Verfaillie (2006) found

large differences in the bias for different actions. They found that the presence of a facing bias was neither related to semantic effects nor to whether the action was familiar or not. A gender bias was reported Brooks et al. (2008) but not confirmed (Schouten et al., 2010). Manera et al. (2012) found that the facing bias for walking disappears when the observer walks on a treadmill, but it was unclear whether this was due to the actor and observer performing the same action or to the observer's head movements.

Our results point more to a low-level explanation. A strong bias was present for the position-scrambled stimulus that did not present biological motion. In contrast, the facing bias was absent in the frame-scrambled stimulus although subjects did recognize the *human form* in these stimuli. We see two implications of these data.

First, the facing bias was absent if the points did not move. The only condition without apparent motion, frame-scrambled, was the only one without a facing bias. This finding was replicated in the main and the control experiments. This absence cannot be explained by a bad recognition of the stimulus because it is known that human observers interpret the frame-scrambled stimulus as human walking (Lange & Lappe, 2007). Also, in the main experiment the subjects' performance, was as good as for the rubber' condition (Figure 5B).

Second, the bias did not depend on whether a human figure was present. The facing-the-observer bias was present in the position-scrambled stimulus although no human walking can be recognized from such a scrambled stimulus. Moreover, the largest bias was found for the rubber stimulus. This stimulus is unnaturally deformed, as the joint angles bend in impossible directions and the limbs are nonrigid.

We therefore argue that the facing-the-observer bias is not related to biological motion mechanisms. Instead one could speculate about different mechanisms. For example, the bias might be related to the bias for looming patterns as opposed to compressing patterns (Ball & Sekuler, 1980; Fahle & Wehrhahn, 1991; Georgeson & Harris, 1978). Such a biased system will respond more with "looming" than with "compressing" to a stimulus of back and forth oscillating dots. Thus, integrating over the biased looming and compressing signals, such a stimulus will appear to be approaching the observer regardless whether it is human or not (Albright, 1989; Lappe & Rauschecker, 1995; Rauschecker, von Grünau, & Poulin, 1987). This explanation would be consistent with the mixed results that have been found for various actions. Vanrie and Verfaillie (2006) found that presentations with little lateral oscillations of the dots resulted in little or no facing bias.

Impression of depth

If the stimulus looks essentially two-dimensional, the observer will tend to underestimate how much the stimulus faces out of the plane of the display (up to a certain degree when the displayed facing direction is obviously close to frontal). As implied by the large error bars in Figure 5A, some subjects generally had a good impression of depth, resulting in a high accuracy, whereas others saw little depth in the stimuli, leading to a systematic negative mean error.

The number of responses along the cardinal axes is more suitable to compare the impression of depth between the six conditions. Such errors were rare for the three conditions with correct *pendular motion* (cf. Table 1), but frequent for the other three conditions. This suggests that pendular motion of the points with respect to each other is an important cue to convey depth. This is in accordance with earlier suggestions (Hoffman & Flinchbaugh, 1982; Johansson, 1973; Kuhlmann et al., 2009). The tendency to respond along the two cardinal axes is also known for the kinetic depth effect and is called the *repulsion effect* (Hiris & Blake, 1996).

Our results are consistent with a kinetic depth effect in combination with an oblique repulsion. Nevertheless, biological motion is a special case in this respect because the kinetic depth is typically defined for rigid objects; biological motion is nonrigid. The kinetic depth in this case can only function if the observer uses implicit knowledge of the body structure and the motions that the human body is capable of.

Cues to depth

The conditions of the main experiment were designed to systematically change the stimulus in different ways to remove some information while leaving other information intact (cf. Table 1).

The human form was intact in four conditions: normal, reversed, marionette, and frame-scrambled. The rubber conditions were physically impossible because the knees and elbows became overextended and the limb segments stretched and compressed during the movement. Only the normal stimuli presented dynamically correct human movements because under normal gravity conditions the marionette movements cannot be performed, and forward walking played in reverse temporal order differs from real backward walking (even if the marionette and reversed stimuli are kinematically and biomechanically valid). Intact *pendular relative movements* of pairs of point-lights were present in the normal, reversed, and the marionette stimulus. Finally, intact *point trajectories* were present in the normal, the rubber, and the position-scrambled stimulus, whereas the *local motions* are dynamically

incorrect for reversed stimuli (cf. human movements). Comparisons of accuracy and precision between different conditions thus allow us to estimate the role of the different depth cues.

Accuracy differed widely between subjects, as expressed by the large error bars in Figure 5A. The within-subject ANOVA showed a clear difference only for the position-scrambled stimulus. This indicates that point trajectories and local motions alone give only poor depth information.

The precision of the judgments is shown in Figure 5B. In this respect the subjects were very similar, but there were large differences between the conditions.

The precision was significantly better in the normal and reversed conditions than in all other conditions. This suggests that the human form was an important cue. On the other hand, neither the human movement nor the local motions were important because both were dynamically incorrect in the reversed condition (cf. Table 1).

Comparing the marionette to the normal stimulus reveals a small but significant decrease in precision. The marionette is consistent with the human form at all times but dynamically incorrect with respect to human movement, point trajectories, and local motions. This reiterates the importance of human form and also suggests that point trajectories might be important. However one should take into account that the configuration of the human form in the marionette stimulus is less familiar than in the normal or reversed conditions.

Precision for the frame-scrambled condition was significantly better than for the position-scrambled condition (Figure 5B). This is consistent with the finding that the human form is an important cue. However, the frame-scrambled condition was significantly worse than the normal and reversed conditions, although all three conditions contain veridical human form information. The difference between the frame-scrambled condition and the normal and reversed conditions lies in pendular motion and point trajectories, which are both absent in the frame-scrambled condition (cf. Table 1). The rubber and marionette stimuli help to resolve this issue. Marionette has correct pendular motion but incorrect point trajectories. Rubber has incorrect pendular motion but correct point trajectories. Since the precision was better in marionette than in rubber stimuli the pendular motions must be important for judging the facing in depth.

This is consistent with a recent finding that the presence of points for the feet improves the detection of small angular deviations from a frontal view of walking (Cai, Yang, Chen, & Jiang, 2011; Kuhlmann et al., 2009). Indeed the notion that human movements are typically pendular has been suggested as an important cue for recognizing biological motion long ago (Hoffman & Flinchbaugh, 1982; Johansson, 1973; Webb &

Aggarwal, 1982). In a slightly different manner, this has also been proposed more recently by Neri (2009).

Lastly, even the precision for the position-scrambled condition was marginally better than the expectancy value for giving the same response on each trial (i.e., 26.2°). Thus, the rhythmic, oscillatory movements of the individual points may have provided some information about the simulated facing in depth.

Altogether, our results support two major cues for the judgment of facing in depth. The first is the familiar human form. The importance of the familiarity of body form for judging the orientation in depth of biological motion is consistent with electrophysiological findings in monkeys (Vangeneugden et al., 2011) and psychophysical findings in humans (Bülthoff et al., 1998; Jackson & Blake, 2010; Kuhlmann et al., 2009; Verfaillie & de Graef, 2000). The second is the pendular movement of the limbs (Hoffman & Flinchbaugh, 1982). The body is an articulated structure with inherent pendular motions. Since these motions are not random but highly characteristic these are a potential additional cue. The data give support to the use of pendular motion as an additional cue for judging the facing in depth.

Supplementary material

There is a Quicktime movie for each of the six conditions of the main experiment. Each movie presents, without interruptions, one walking cycle for the facing in depth of 0°, 30°, 60°, 90°, 120°, 150°, and 180°.

Acknowledgments

Commercial relationships: none.
Corresponding author: Marc H. E. de Lussanet.
Email: lussanet@wwu.de.
Address: Psychologisches Institut, Westf. Wilhelms-Universität Münster, Münster, Germany.

References

Albright, T. D. (1989). Centrifugal directionality bias in the middle temporal visual area (MT) of the macaque. *Visual Neuroscience, 2*, 177–188. [PubMed]

Ball, K., & Sekuler, R. (1980). Human vision favors centrifugal motion. *Perception, 9*(3), 317–325. [PubMed]

Beintema, J. A., & Lappe, M. (2002). Perception of biological motion without local image motion. *Proceedings of the National Academy of Sciences of the USA, 99*, 5661–5663, doi:10.1073/pnas.082483699. [PubMed]

Blanz, V., Tarr, M. J., & Bülthoff, H. H. (1999). What object attributes determine canonical views? *Perception, 28*, 575–599.

Brooks, A., Schouten, B., Troje, N. F., Verfaillie, K., Blanke, O., & van der Zwan, R. (2008). Correlated changes in perceptions of the gender and orientation of ambiguous biological motion figures. *Current Biology, 18*(17), R728–R729, doi:10.1016/j.cub.2008.06.054. [PubMed]

Bülthoff, I., Bülthoff, H. H., & Sinha, P. (1998). Top-down influences on stereoscopic depth-perception. *Nature Neuroscience, 1*, 254–257. [PubMed]

Cai, P., Yang, X., Chen, L., & Jiang, Y. (2011). Motion speed modulates walking direction discrimination: The role of the feet in biological motion perception. *Chinese Science Bulletin, 56*(19), 2025–2030, doi:10.1007/s11434-011-4528-6.

Chang, D. H. F., & Troje, N. F. (2009). Acceleration carries the local inversion effect in biological motion perception. *Journal of Vision, 9*(1):19, 1–17, http://journalofvision.org/9/1/19/, doi:10.1167/9.1.19. [PubMed] [Article]

Fahle, M., & Wehrhahn, C. (1991). Motion perception in the peripheral visual field. *Graefe's Archive for Clinical and Experimental Ophthalmology, 229*(5), 430–436, doi:10.1007/BF00166305. [PubMed]

Foster, D. H., & Gilson, S. J. (2002). Recognizing novel three-dimensional objects by summing signals from parts and views. *Proceedings of the Royal Society B: Biological Sciences, 269*, 1939–1947, doi:10.1098/rspb.2002.2119.

Georgeson, M. A., & Harris, M. G. (1978). Apparent foveofugal drift of counterphase gratings. *Perception, 7*, 527–536. [PubMed]

Giese, M. A. (2004). Neural model for biological movement recognition: A neurophysiologically plausible theory. In L. M. Vaina, S. A. Beardsley, & S. Rushton (Eds.), *Optic flow and beyond.* (pp. 443–470). Dordrecht, NL: Kluwer Academic Publishers.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential

de Lussanet & Lappe

processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron, 24*, 187–203.

Harries, M. H., Perrett, D. I., & Lavender, A. (1991). Preferential inspection of views of 3-D model heads. *Perception, 20*(5), 669–680, doi:10.1068/p200669.

Hill, H., & Johnston, A. (2007). The hollow-face illusion: Object-specific knowledge, general assumptions or properties of the stimulus? *Perception, 36*(2), 199–223, doi:10.1068/p5523. [PubMed]

Hiris, E., & Blake, R. (1996). Direction repulsion in motion transparency. *Visual Neuroscience, 13*(1), 187–197, doi:10.1017/S0952523800007227. [PubMed]

Hoffman, D. D., & Flinchbaugh, B. E. (1982). The interpretation of biological motion. *Biological Cybernetics, 42*, 195–204, doi:10.1007/BF00340076. [PubMed]

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*, 65–70.

Jackson, S., & Blake, R. (2010). Neural integration of information specifying human structure from form, motion and depth. *The Journal of Neuroscience, 30*(3), 838–848, doi:10.1523/jneurosci.3116-09.2010. [PubMed]

Jackson, S., Cummins, F., & Brady, N. (2008). Rapid perceptual switching of a reversible biological figure. *PLoS ONE, 3*(12), e3982, doi:10.1371/journal.pone.0003982.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*, 201–211, doi:10.3758/BF03212378.

Jokisch, D., & Troje, N. F. (2003). Biological motion as a cue for the perception of size. *Journal of Vision, 3*(4):1, 252–264, http://journalofvision.org/3/4/1/, doi:10.1167/3.4.1. [PubMed] [Article]

Kuhlmann, S., de Lussanet, M. H. E., & Lappe, M. (2009). Perception of limited lifetime biological motion from different viewpoints. *Journal of Vision, 9*(10):11, 1–14, http://journalofvision.org/9/10/11, doi:10.1167/9.10.11. [PubMed] [Article]

Lange, J., Georg, K., & Lappe, M. (2006). Visual perception of biological motion by form: A template-matching analysis. *Journal of Vision, 6*(8):6, 836–849, http://journalofvision.org/6/8/6, doi:10.1167/6.8.6. [PubMed] [Article]

Lange, J., & Lappe, M. (2006). A model of biological motion perception from configural form cues. *The Journal of Neuroscience, 26*(11), 2894–2906, doi:10.1523/jneurosci.4915-05.2006. [PubMed]

Lange, J., & Lappe, M. (2007). The role of spatial and temporal information in biological motion perception. *Advances in Cognitive Psychology, 3*(4), 419–429. [PubMed]

Lappe, M., & Rauschecker, J. P. (1995). Motion anisotropies and heading detection. *Biological Cybernetics, 72*, 261–277. [PubMed]

Lee, J., & Wong, W. (2004). A stochastic model for the detection of coherent motion. *Biological Cybernetics, 91*, 306–314, doi:10.1007/s00422-004-0516-0. [PubMed]

Manera, V., Cavallo, A., Chiavarino, C., Schouten, B., Verfaillie, K., & Becchio, C. (2012). Are you approaching me? Motor execution influences perceived action orientation. *PLoS ONE, 7*(5):e37514, doi:10.1371/journal.pone.0037514.

Michels, L., Kleiser, R., de Lussanet, M. H. E., Seitz, R. J., & Lappe, M. (2009). Brain activity for peripheral biological motion in the posterior superior temporal gyrus and the fusiform gyrus: Dependence on visual hemifield and view orientation. *NeuroImage, 45*, 151–159, doi:10.1016/j.neuroimage.2008.10.063.

Neri, P. (2009). Wholes and subparts in visual processing of human agency. *Proceedings of the Royal Society B: Biological Sciences, 276*(1658), 861–869, doi:10.1098/rspb.2008.1363. [PubMed]

Peissig, J. J., & Tarr, M. J. (2006). Visual object recognition: Do we know more now than we did 20 years ago? *Annual Review of Psychology, 58*(1), 75–96, doi:10.1146/annurev.psych.58.102904.190114.

Perrett, D. I., & Harries, M. H. (1988). Characteristic views and the visual inspection of simple faceted and smooth objects: 'Tetrahedra and potatoes'. *Perception, 17*(6), 703–720, doi:10.1068/p170703. [PubMed]

Rauschecker, J. P., von Grünau, M. W., & Poulin, C. (1987). Centrifugal organization of direction preferences in the cat's lateral suprasylvian visual cortex and its relation to flow field processing. *The Journal of Neuroscience, 7*(4), 943–958. [PubMed]

Schouten, B., Troje, N. F., Brooks, A., van der Zwan, R., & Verfaillie, K. (2010). The facing bias in biological motion perception: Effects of stimulus gender and observer sex. *Attention, Perception, & Psychophysics, 72*(5), 1256–1260, doi:10.3758/APP.72.5.1256. [PubMed]

Schouten, B., Troje, N. F., & Verfaillie, K. (2011). The facing bias in biological motion perception: Structure, kinematics, and body parts. *Attention, Perception, & Psychophysics, 73*(1), 130–143, doi:10.3758/s13414-010-0018-1. [PubMed]

Schouten, B., Troje, N. F., Vroomen, J., & Verfaillie, K. (2011). The effect of looming and receding sounds on the perceived in-depth orientation of depth-ambiguous biological motion figures. *PLoS ONE, 6*(2), e14725, doi:10.1371/journal.pone. 0014725.

Sweeny, T. D., Haroz, S., & Whitney, D. (2012). Reference repulsion in the categorical perception of biological motion. *Vision Research, 64*, 26–34, doi: 10.1016/j.visres.2012.05.008. [PubMed]

Troje, N. F., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research, 36*(12), 1761–1771, doi:10.1016/0042-6989(95)00230-8.

Troje, N. F., & Westhoff, C. (2006). The inversion effect in biological motion perception: Evidence for a "life detector"? *Current Biology, 16*, 821–824, doi: 10.1016/j.cub.2006.03.022.

Vangeneugden, J., De Maziere, P. A., Van Hulle, M. M., Jaeggli, T., Van Gool, L., & Vogels, R. (2011). Distinct mechanisms for coding of visual actions in macaque temporal cortex. *The Journal of Neuroscience, 31*(2), 385–401, doi:10.1523/jneurosci. 2703-10.2011.

Vanrie, J., Dekeyser, M., & Verfaillie, K. (2004). Bistability and biasing effects in the perception of ambiguous point-light walkers. *Perception, 33*, 547–560. [PubMed]

Vanrie, J., & Verfaillie, K. (2006). Perceiving depth in point-light actions. *Perception & Psychophysics, 68*(4), 601–612, doi:10.3758/BF03208762. [PubMed]

Verfaillie, K., & de Graef, P. (2000). Transsaccadic memory for position and orientation of saccade source and target. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1243–1259. [PubMed]

Wang, G., Tanifuji, M., & Tanaka, K. (1998). Fuctional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging. *Neuroscience Research, 32*, 33–46. [PubMed]

Watson, T. L., Johnston, A., Hill, H. C. H., & Troje, N. F. (2005). Motion as a cue for viewpoint invariance. *Visual Cognition, 12*(7), 1291–1308.

Webb, J. A., & Aggarwal, J. K. (1982). Structure from motion of rigid and jointed objects. *Artificial Intelligence, 19*, 107–130, doi:10.1.1.77.6863.

Yellott, J. I., & Kaiwi, J. L. (1979). Depth inversion despite stereopsis: The appearance of random-dot stereograms on surfaces seen in reverse perspective. *Perception, 8*(2), 135–142. [PubMed]