

Frames of reference for biological motion and face perception

Dorita H. F. Chang

Centre for Neuroscience Studies, Queen's University,
Kingston, Canada



Laurence R. Harris

Department of Psychology and Centre for Vision Research,
York University, Toronto, Canada



Nikolaus F. Troje

Department of Psychology, School of Computing, and
Centre for Neuroscience Studies, Queen's University,
Kingston, Canada, &
Centre for Vision Research, York University, Toronto,
Canada



We investigated the roles of egocentric, gravitational, and visual environmental reference frames for face and biological motion perception. We tested observers on face and biological motion tasks while orienting the visual environment and the observer independently with respect to gravity using the York Tumbling Room. The relative contribution of each reference frame was assessed by arranging pairs of frames to be either aligned or opposed to each other while rendering the third uninformative by orienting it sideways relative to the stimulus. The perception of both biological motion and faces were optimal when the stimulus was aligned with egocentric coordinates. However, when the egocentric reference frame was rendered uninformative, the perception of biological motion, but not faces, relied more on stimulus alignment with gravity rather than the visual environment.

Keywords: face perception, biological motion, reference frames

Citation: Chang, D. H. F., Harris, L. R., & Troje, N. F. (2010). Frames of reference for biological motion and face perception. *Journal of Vision*, 10(6):22, 1–11, <http://www.journalofvision.org/content/10/6/22>, doi:10.1167/10.6.22.

Introduction

Many objects become unfamiliar or even unrecognizable when they are seen in other than their veridical orientation (e.g., Sumi, 1984; Thompson, 1980). As noted by Rock (1973), this observation contradicts the early Gestalt theorists who considered that the essential information for the perception of form was contained primarily in the geometrical relationship of the features—that is, perceived form would remain unchanged as long as these internal relationships were maintained. Orientation specificity demonstrates that the stimuli are encoded not just in terms of internal relations but also in relation to external reference frames. Such reference frames can be either egocentric or allocentric (see Howard, 1982 for a review). Egocentric reference frames include the retina, head, and body. Allocentric reference frames include gravity and the visual environment. The reference system that is most important in determining the way in which we perceive an object seems to depend on the stimulus class involved. Here, we consider the perception of faces and biological motion.

The recognition of a face is impaired if it is inverted (see Valentine, 1988 for a review). The inversion effect

observed for face perception is due primarily to disruptions of configural rather than featural processing (Freire, Lee, & Symons, 2000). Inversion effects have also been reported for the perception of biological motion (e.g., Sumi, 1984). Biological motion appears to carry at least two distinct inversion effects (Troje & Westhoff, 2006). While there appears to be an inversion effect that is due to impaired processing of the global configural shape of the walker as conveyed by the display's spatiotemporal organization (e.g., Bertenthal & Pinto, 1994), there is also a second inversion effect that is associated with local motion signals of the distal limbs (Chang & Troje, 2009a; Shipley, 2003; Troje & Westhoff, 2006). To this end, Troje and Westhoff (2006) presented observers with intact and scrambled walker displays in which the dot's spatial organization was perturbed. Significantly, observers could retrieve the facing direction of the walker not only for the coherent displays but also for the scrambled displays, which retained solely local motion information. Moreover, an inversion effect was observed for both intact and scrambled displays. The authors later showed that the cues to direction in the scrambled displays and the associated inversion effect were carried entirely by the local motion of the feet. Subsequent work has shown that the local inversion effect depends on vertical acceleration contained

in the foot motion (Chang & Troje, 2009a). Moreover, the mechanisms underlying the perception of global motion-mediated structure and local motion signals are dissociable according to a variety of behavioral characteristics such as sensitivity to masking or susceptibility to learning (Chang & Troje, 2009b).

Few studies have investigated the reference frames in which animate motions and faces are coded. Observers use implicit knowledge about the direction and effects of gravity when interpreting biological and inanimate events (e.g., Jokisch & Troje, 2003; Pittenger, 1985; Runeson & Frykholm, 1981; Shipley, 2003; Stappers & Waller, 1993). Making assumptions with regard to the direction of gravity, however, does not necessarily implicate direct measurements of gravitational acceleration (e.g., via input from the vestibular system). The visual system may simply take advantage of the fact that gravity is typically aligned with egocentric coordinates. Still, there is reason to believe that an allocentric system may be involved for the perception of dynamic events (Bingham, Schmidt, & Rosenblum, 1995; Indovina et al., 2005; Lopez, Bachofner, Mercier, & Blanke, 2009). Bingham et al. (1995) found that the recognition of point-light-defined events was stronger for displays that were upright rather than inverted with respect to gravity, regardless of the observer's orientation in space and concluded that point-light events are perceived in relation to a gravitational rather than an egocentric frame of reference. Later findings, however, appear to be inconsistent with this conclusion. Troje (2003) found that performance on a biological motion task depended only on whether the display was aligned with the observer regardless of the observer's orientation in space, suggesting egocentric coding. A study with infants also suggested that the egocentric reference frame dominates in the coding of animate motions (Kushiro, Taga, & Watanabe, 2007).

In the domain of face perception, early evidence suggested that the egocentric system is also the dominant frame of reference. Kohler (1940) and Rock (1988) reported that the recognition of faces presented upright with respect to gravity and the visual environment was impaired for observers with their heads held upside down. The finding of Troje (2003) that performance on a face recognition task depended only on stimulus alignment with the observer corroborates these early reports. However, recent findings by Lobmaier and Mast (2007) seem to suggest a role for gravity as a reference frame for the coding of faces.

To our knowledge, no studies thus far have provided a clear experimental distinction between the roles of gravitational and visual environmental reference frames for the perception of biological motion and faces. Moreover, whether the global and local aspects of biological motion are coded in the same reference systems is unknown. In the present study, we teased apart the contributions of three reference frames (egocentric, visual environment, and gravity) by placing observers inside the York University “tumbling room”—a room furnished with strong directional visual cues (e.g., table, chair, drapery) that can be rotated about a horizontal axis by 360 degrees (Figure 1A).

Inside the room (Figure 1B), an observer can also be rotated (rolled) independently around the same axis. Using this facility, we investigated the perception of biological motion and faces by creating configurations in which two reference frames were put into conflict (aligned with or opposed to the stimulus) while the third was rendered uninformative by arranging it to be orthogonal to the stimulus. An entirely balanced design resulted in 12 experimental configurations illustrated in Figure 2. If biological motion and faces are largely coded by egocentric (we do not distinguish between the retina, head, or rest of the

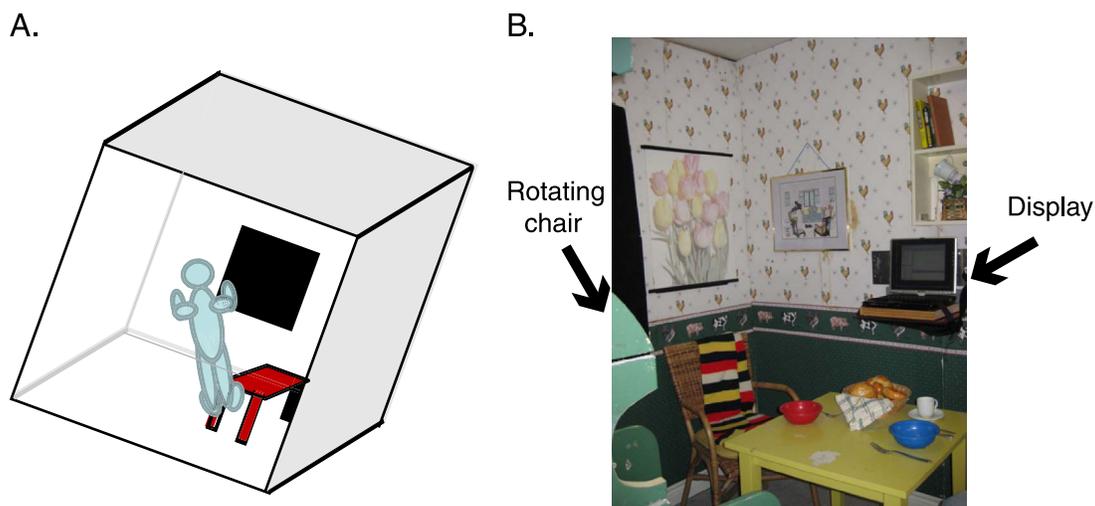


Figure 1. (A) Schematic depiction of the “tumbling room.” The room and observer can be rotated independently by 360 degrees about a horizontal axis. (B) The interior of the tumbling room. The display, located across from the observer's chair, is mounted at the axis of rotation.

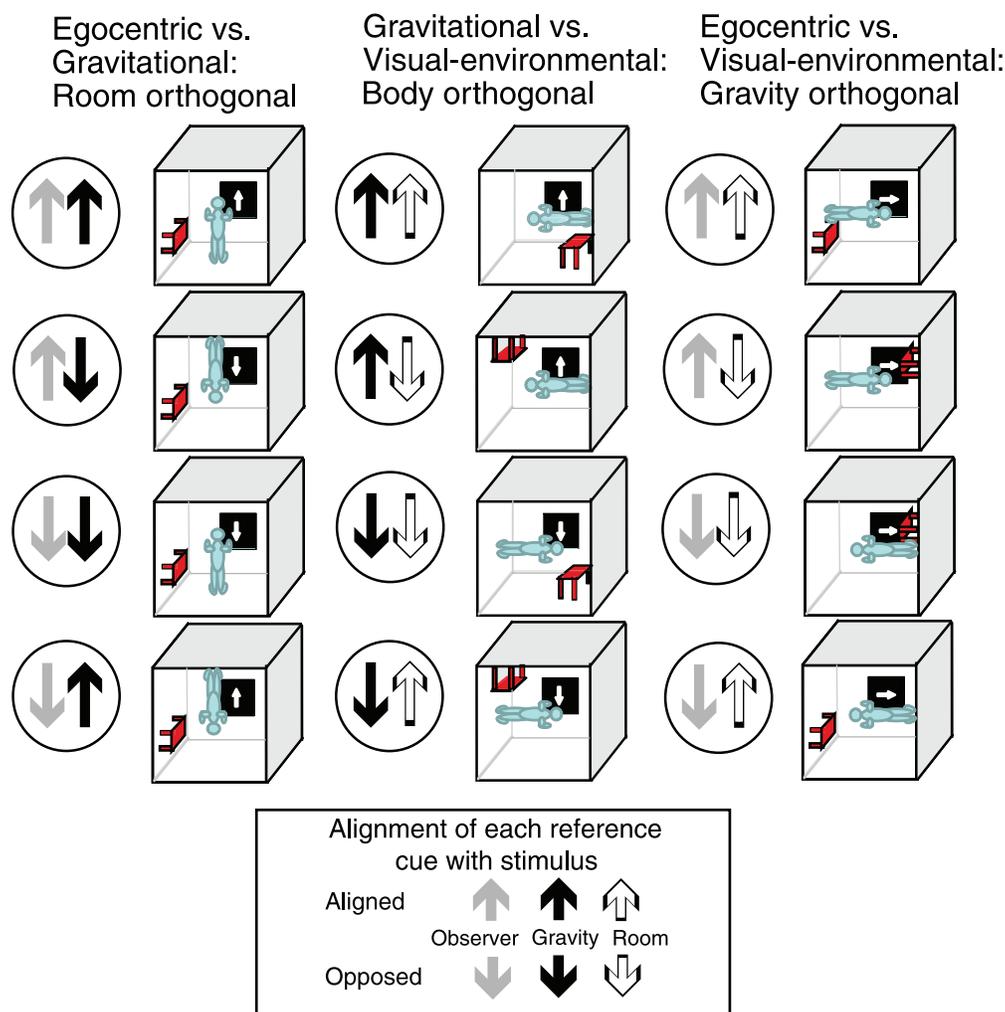


Figure 2. Schematic depictions of the 12 room/observer/stimulus configurations. The white arrow on the display represents the orientation of the stimulus where an upward arrow is upright and a downward arrow is inverted with respect to gravity. For each configuration, the third reference frame was rendered uninformative by orienting it orthogonal to the stimulus orientation. The large arrows describe the reference frames in relation to the stimulus. Upward and downward arrows indicate that the stimulus was upright or inverted with respect to the observer (gray), gravity (black), or the room (white), respectively.

body here), gravitational, or visual environmental coordinates, performance should be best when the stimulus is aligned with the respective reference frame.

We investigated the perception of both global motion-mediated structure in biological motion and local motion by manipulating the organization of the walker and the type of mask during a biological motion direction discrimination task. Specifically, the perception of global motion-mediated structure was addressed by placing veridical walkers inside a mask of additional walker dots moving in the opposing direction. This manipulation equated the local motion of the display. Consequently, the task could only be solved by retrieving the global form of the walker. The local aspect of biological motion was addressed by placing walkers that had their individual motion trajectories spatially perturbed (thereby destroying global structure) inside a mask of stationary flickering

dots. These displays could thus only be solved based upon local motion cues. Additionally, we investigated reference frames for face perception via a same-different face recognition task previously employed by Troje (2003).

Methods

Participants

Twelve naive observers, 19–63 years of age (mean age of 34 years; 7 males, 5 females), recruited from the Centre for Vision Research at York University, participated in this experiment. All observers received monetary compensation for their time and had normal or corrected-to-

normal vision. The procedures of this experiment were approved by both the York University Ethics Review Board and the Queen's University General Research Ethics Board.

Stimuli and apparatus

The orientation of egocentric and visual environmental cues with respect to gravity was manipulated using the York tumbling room (Figure 1). Both the room ($2.4 \text{ m} \times 2.4 \text{ m} \times 2.4 \text{ m}$) and the observer's chair could be rotated independently about a common axis that was normal to the fronto-parallel plane of the observer. Stimuli were generated using MATLAB (Mathworks, Natick, MA) with extensions from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) and displayed on a laptop equipped with a 12-inch screen mounted on the wall of the tumbling room that faced the observer and was centered on the axis of rotation. The observer was strapped securely to the chair and had a wireless mouse strapped to one leg. Stimuli were viewed binocularly at a distance of 147 cm.

Biological motion stimuli

The biological motion stimuli were derived from point-light sequences of a stationary-walking human, cat, and pigeon (Figure 3). The human walker was computed as the average walker from motion-captured data of 50 men and 50 women (Troje, 2002) and was represented by a set of 11 dots. The cat sequence was created by sampling 14 points from single frames of a video sequence showing a cat walking on a treadmill. The pigeon sequence was created from motion-captured data obtained from a pigeon fitted with 11 markers. All walkers were presented in sagittal view (facing either rightward or leftward). All

sequences were shown at their veridical gait frequencies of 0.93 Hz, 1.7 Hz, and 1.6 Hz for the human, cat, and pigeon, respectively. On each trial, the starting position of the walker within its gait cycle was selected randomly.

Walkers were presented coherently (Figure 3A) or spatially scrambled (Figure 3B). Scrambled walkers were created by displacing each point's trajectory to a randomly selected position within the display area. Coherent walkers were embedded in a scrambled walker mask. This mask was comprised of dots carrying veridical trajectories of the walker displaced randomly on the screen and moving in a direction opposite to that of the target walker. The number of mask dots was set to be identical to the number of dots comprising the target walker in order to equate the directionality of the individual trajectories of all dots contained in the display. These coherent-walker-and-scrambled-mask displays thus retained solely global structure-from-motion information. Scrambled walkers were embedded in a flicker mask comprised of randomly positioned stationary dots with a limited lifetime of 125 ms. Twice the number of dots comprising the walker was used for the mask. These scrambled-walker-and-flicker-mask displays retained solely local motion information.

The walkers subtended visual angles of $2.8 \times 5.2 \text{ deg}$, $5.2 \times 2.8 \text{ deg}$, and $4.0 \times 4.0 \text{ deg}$ for the human, cat, and pigeon, respectively. The masking dots were contained in an area measuring $5.9 \times 5.9 \text{ deg}$ at the center of the screen.

Face stimuli

The face stimuli were images of human faces ($n = 30$) derived from 3D models of laser-scanned heads (Troje & Bühlhoff, 1996). The faces were shown turned either 5 deg or 10 deg to the left or right of frontal view. Hair was removed digitally and the images were rendered such that only the head and the upper section of the neck were visible (Figure 3C). Half of the faces used were female. The faces subtended visual angles of $4.4 \times 5.6 \text{ deg}$ on average.

Procedure

Two sequences of the twelve configurations shown in Figure 2 were run in two sessions separated by at least 1 day. During one session, observers completed a face recognition task and during the other session they completed a biological motion direction discrimination task. The order of the two tasks was counterbalanced among participants.

For each configuration, subjects completed a block of trials that lasted approximately 2 min. The next configuration was then selected and the room and observer (with eyes shut) were rotated at constant velocity to the new orientation. When the chair was moved, the route that minimized time spent with the observer upside down was

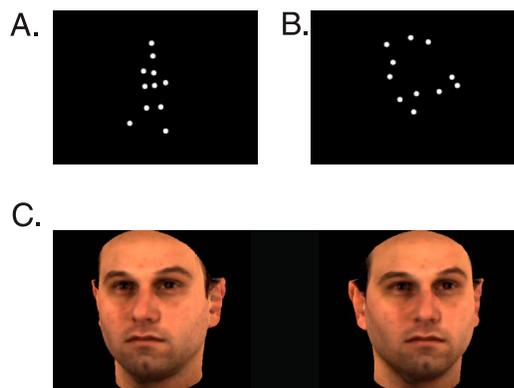


Figure 3. Static frames taken from point-light sequences used in the biological motion task depicting a (A) coherent human walker and (B) scrambled human walker in which the constituent dots are spatially displaced. (C) For the face recognition task, faces were shown turned either 5 deg or 10 deg in profile to the left or right of frontal view.

chosen. The order in which the configurations were run was randomized with the constraint that no two configurations involving an observer held upside down were run consecutively.

Biological motion task

On each trial, a single walker was presented at the center of the screen and the observer was required to indicate whether the walker was facing to the right or left (relative to them). When the stimulus was rendered sideways with respect to the observer, the task was to indicate whether the walker was facing upward or downward. Instructions were given for each of the 12 configurations before each block of trials commenced. Each stimulus was presented on the screen for 1000 ms after which it was removed and replaced with a prompt (left/right, up/down arrows) on the screen until a response was provided.

Participants completed a practice block of trials with the orientation of the room, chair, and gravity all aligned in the upright position. The practice block consisted of 24 trials and was comprised of all possible combinations of left and right facing, coherent, and scrambled versions of the human, cat, or pigeon presented upright or inverted with respect to each reference frame. For the practice trials only, the walkers were shown unmasked.

After the practice trials had been completed, the participant and room were rotated into the first configuration. The participants were instructed to initiate each test block by using the mouse and to call out to the experimenter who was standing outside the tumbling room when the block was over. For the test trials, coherent walkers were embedded in a scrambled walker mask (global cues) while scrambled walkers were embedded in the flicker mask (local cues). Each test block consisted of 36 trials and was comprised of all possible combinations of left and right facing, coherent, and scrambled versions of the human, cat, and pigeon (each repeated three times) presented in random order. A total of 432 trials were completed across all configurations. Feedback was not given.

Face task

On each trial in this task, two faces were presented sequentially and the observer was required to indicate whether the faces were of the same or different person. Each face was presented at the center of the screen for 100 ms and the two faces were separated by a 500-ms inter-stimulus interval. Following the disappearance of the face in the second interval, a prompt appeared and remained until a response was provided.

The face session began with a practice block comprised of 16 trials, consisting of male and female faces presented in upright or inverted orientations (with respect to each reference frame). The test blocks then followed. On all trials, the gender of the two faces was always identical. Half of the trials featured male faces while the other half featured female faces. On each trial, one face was presented

facing 5 deg in one direction from a frontal view and the other was presented facing 10 deg in the opposite direction. Half of the trials presented the same individuals while the other half presented different individuals. Each test block consisted of 48 trials. A total of 576 trials were completed across all configurations. Feedback was not given.

Results

Global structure from biological motion

General performance and stimulus specificity

This section of the results refers to all trials showing coherent walkers in a scrambled walker mask requiring discriminations based upon the retrieval of global structure from motion. The data, collapsed across all 12 configurations with reference frames aligned, opposed, or

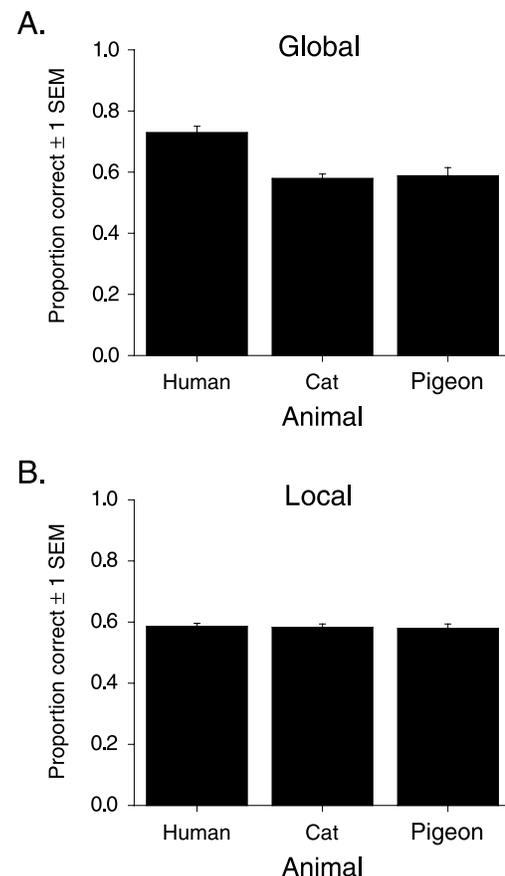


Figure 4. Direction discrimination accuracies for the biological motion task, expressed in terms of the proportions of correct responses. The data for displays involving (A) global structure and (B) local motion cues for the three walker types and collapsed across all configurations (blocks) in which the reference frames are aligned, opposed, or orthogonal to the stimulus as shown in Figure 2. Error bars represent ± 1 standard error of the mean.

orthogonal to the stimulus, are presented in Figure 4A. Overall performance is relatively poor as some of the configurations resulted in chance-level performance (Figure 5). A one-way ANOVA revealed a significant main effect of animal type, $F(2, 22) = 14.34, p = 0.003$. Post-hoc Tukey’s comparisons of the three animal types

revealed that performance was better for the human walker than for the cat ($p = 0.002$) and pigeon ($p = 0.004$), which did not differ ($p > 0.500$).

This observation confirms earlier observations (Chang & Troje, 2009b) but is not directly relevant to understanding the contributions of reference systems. The

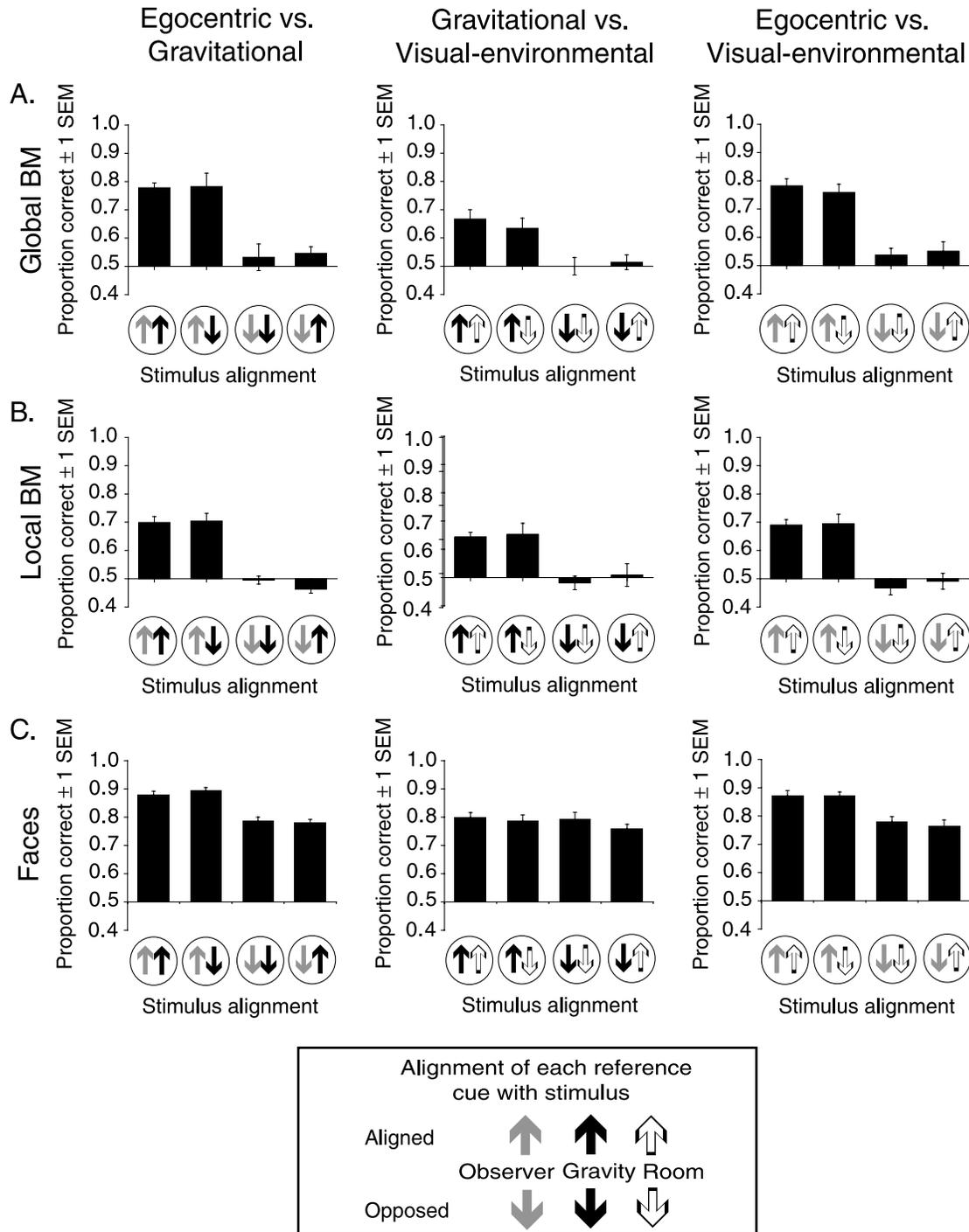


Figure 5. Proportions of correct responses for (A) global biological motion discrimination, (B) local biological motion discrimination, and (C) face recognition across the four blocks corresponding to comparisons between (left) the egocentric and gravitational coordinates, (middle) the gravitational and visual environmental coordinates, and (right) the egocentric and visual environmental coordinates. The arrows below the x-axes depict stimulus alignment with the relevant reference frames. Error bars represent \pm standard error of the mean.

different walkers were used to show that our results are not specific to a human structure and to introduce some variability in order to render it less likely that observers would use stimulus-specific artifacts to solve the task. For this reason, and to facilitate later comparisons between the different tasks, we subsequently pooled the data over walker types and examined them across the separate subsets of four blocks, constituting comparisons between pairs of reference frames.

Egocentric vs. gravitational reference frames

The data for the blocks corresponding to a comparison between egocentric and gravitational frames of reference for the global-structure task are presented in the first panel of [Figure 5A](#). A 2 (stimulus upright/inverted with respect to observer) \times 2 (stimulus upright/inverted with respect to gravity) ANOVA indicated a significant main effect of stimulus alignment with the observer, $F(1, 11) = 27.34$, $p < 0.001$, reflecting the fact that accuracies were highest when the stimulus was aligned with the observer. There was no effect of stimulus alignment with gravity. Additional t -tests indicated accuracy rates for the two conditions in which the stimulus was inverted with respect to the egocentric system were not significantly different from chance level ($p > 0.06$ for both).

Gravitational vs. visual environmental reference frames

The second panel of [Figure 5A](#) shows the data corresponding to a comparison between the gravitational and visual environmental reference frames for the global-structure subtask. A two-way ANOVA indicated a significant main effect of gravity, $F(1, 11) = 13.17$, $p = 0.004$. Performance was best when the stimulus was upright with respect to gravity. There was no effect of stimulus alignment with the room. Additional t -tests confirmed that accuracy rates for the two blocks in which the stimulus was upright with respect to gravity were in fact above chance level ($p < 0.003$ for both). By contrast, accuracy rates for the two blocks in which the stimulus was inverted with respect to gravity were not significantly different from chance level ($p > 0.5$ for both).

Egocentric vs. visual environmental reference frames

The data corresponding to the blocks comparing egocentric and visual environmental frames of reference for the global-structure subtask are presented in the third panel of [Figure 5A](#). A two-way ANOVA showed a significant main effect of stimulus alignment with the observer, $F(1, 11) = 48.46$, $p < 0.001$. Performance was highest when the stimulus was aligned with the observer. There was no effect of stimulus alignment with the room. Additional t -tests indicated that accuracy rates for the two blocks in which the stimulus was inverted with respect to

the egocentric system were not significantly different from chance level ($p > 0.15$ for both).

Local biological motion

General performance and stimulus specificity

This section refers to all trials showing scrambled walkers masked with flickering dots—stimuli containing solely local cues to indicate facing direction. The data for the three walker types, collapsed across all blocks, are presented in [Figure 4B](#). A one-way ANOVA showed no difference in performance rates across the various animals.

Egocentric vs. gravitational reference frames

The first panel of [Figure 5B](#) presents the data for the four blocks corresponding to a comparison between egocentric and gravitational frames of reference for the local subtask. A two-way ANOVA indicated a significant main effect of stimulus alignment with the observer, $F(1, 11) = 88.01$, $p < 0.001$. Accuracy was highest when the stimulus was aligned with the observer. Stimulus alignment with gravity did not affect performance. Additional t -tests indicated accuracy rates for the two conditions in which the stimulus was inverted with respect to the egocentric system were not significantly different from chance level ($p > 0.05$ for both).

Gravitational vs. visual environmental reference frames

The data corresponding to a comparison between the gravitational and visual environmental reference frames for the local subtask are shown in the second panel of [Figure 5B](#). A two-way ANOVA for these blocks revealed a significant main effect of gravity, $F(1, 11) = 18.58$, $p = 0.001$. Performance was best when the stimulus was upright with respect to gravity. There was no effect of stimulus alignment with the room. Additional t -tests confirmed that accuracy rates for the two blocks in which the stimulus was upright with respect to gravity were in fact above chance level ($p < 0.004$ for both). By contrast, accuracy rates for the two blocks in which the stimulus was inverted with respect to gravity were not significantly different from chance level ($p > 0.5$ for both).

Egocentric vs. visual environmental reference frames

The third panel of [Figure 5B](#) presents the data corresponding to the blocks comparing egocentric and visual environmental frames of reference for the local subtask. Here, a two-way ANOVA showed a significant main effect of stimulus alignment with the observer, $F(1, 11) = 33.64$, $p < 0.001$, reflecting the fact that performance was best when the stimulus was aligned with the observer. Again, there was no effect of stimulus

orientation with respect to the room. Additional *t*-tests indicated that accuracy rates for the two blocks in which the stimulus was inverted with respect to the egocentric system were not significantly different from chance level ($p > 0.20$ for both).

Face recognition

Overall performance

The data for the face task, expressed in terms of proportions of correct responses, were first analyzed by means of a paired *t*-test comparing the two genders of the face stimuli. The analysis indicated that performances for the male and female faces did not differ. Therefore, the data were collapsed for further analyses.

Egocentric vs. gravitational reference frames

The first panel of [Figure 5C](#) shows comparisons between the egocentric and gravitational frames of reference. A two-way ANOVA indicated a significant main effect of stimulus alignment with the observer, $F(1, 11) = 160.33$, $p < 0.001$. Performance on the face task was best in conditions where the stimulus was aligned with the observer rather than with gravity.

Gravitational vs. visual environmental reference frames

The data corresponding to comparisons between the gravitational and visual environmental frames of reference for the face task are presented in the second panel of [Figure 5C](#). Here, a two-way ANOVA showed no effects of gravity or the room.

Egocentric vs. visual environmental reference frames

The data corresponding to comparisons between the egocentric and visual environmental frames of reference are presented in third panel of [Figure 5C](#). A two-way ANOVA indicated a significant main effect of stimulus orientation with respect to the observer, $F(1, 11) = 107.41$, $p < 0.001$, reflecting the fact that accuracies were highest when the stimulus was aligned with the observer rather than the room.

Discussion

Biological motion and faces are primarily egocentrically coded

We found no significant differences in terms of the relevance of the various reference frames for discriminating

the direction of the global and local biological motion stimuli. These two types of displays were designed such that different cues must be exploited to retrieve the facing direction of the walker. The “global” displays could be solved only by retrieving the coherent structure of the figure. Orientation effects observed for these particular displays must be due to impaired processing of this structure-from-motion information. By contrast, the “local” displays were devoid of structure and contained solely local cues to direction. The inversion effect observed for the perception of local biological motion has been shown to be carried by the acceleration contained in the foot motion (Chang & Troje, 2009a). The similarity of the data for the two types of displays suggests that both types of information are coded in a similar manner.

In line with earlier work (Troje, 2003), our results confirm that both biological motion and face perception are dominated by an egocentric frame of reference. For all types of stimuli, performance was best when the stimulus was aligned with the egocentric reference frame, regardless of the orientation of the observer with respect to gravity. Performance dropped significantly when the stimulus was inverted with respect to the observer whether or not it was aligned with gravity or with the room.

Of particular interest are the results for the conditions in which the egocentric reference frame was rendered uninformative. For these conditions, the results obtained for both biological motion subtasks were different from those obtained for the face recognition task. Performance on the biological motion tasks (both for conditions requiring global structure-from-motion or local discriminations) was higher when the stimulus was aligned with gravity rather than the room suggesting contribution from a gravity-based reference frame. This was not the case for the face recognition task.

We note that although the biological motion and face recognition tasks used in this study differed in terms of their overall difficulty, it is unlikely that this difference can account for the differential effects of the gravity reference frame observed here. An inspection of the relevant comparisons of interest revealed that overall performance for the face recognition task ([Figure 5C](#), middle panel) was significantly worse than for conditions in which the stimuli were aligned with the egocentric frame of reference ([Figure 5C](#), left and right panels, first two data bars). This suggests that the lack of an effect of the gravity reference frame for the face recognition task is not due to mere performance saturation. Nonetheless, it should be acknowledged that the generalizability of these findings to other tasks, and more specifically, whether an effect of the gravity reference frame may be found for face perception using other tasks is unknown.

The data for both tasks are summarized in [Figure 6A](#). For the face recognition task, it is clear that the data best fit the prediction made by the assumption that egocentric coordinates dominate perception. However, for the two biological motion tasks, the data show an additional

influence of stimulus orientation with respect to gravity. We therefore modeled the contribution of all three reference systems.

Modeling the data with a linear model

We fit the data with a simple linear model by modeling performance rate, $r = x + eE + gG + vV$, where e , g , and v

were weightings for contributions from the egocentric, gravitational, and visual environmental reference frames, respectively, added to base performance x . The variables E , G , and V assume values of 1, -1, or 0 if the stimulus were aligned, opposed, or orthogonal to the reference frame, respectively.

The model was fit separately to the global biological motion, local biological motion, and face recognition data. The relative weightings obtained are summarized as

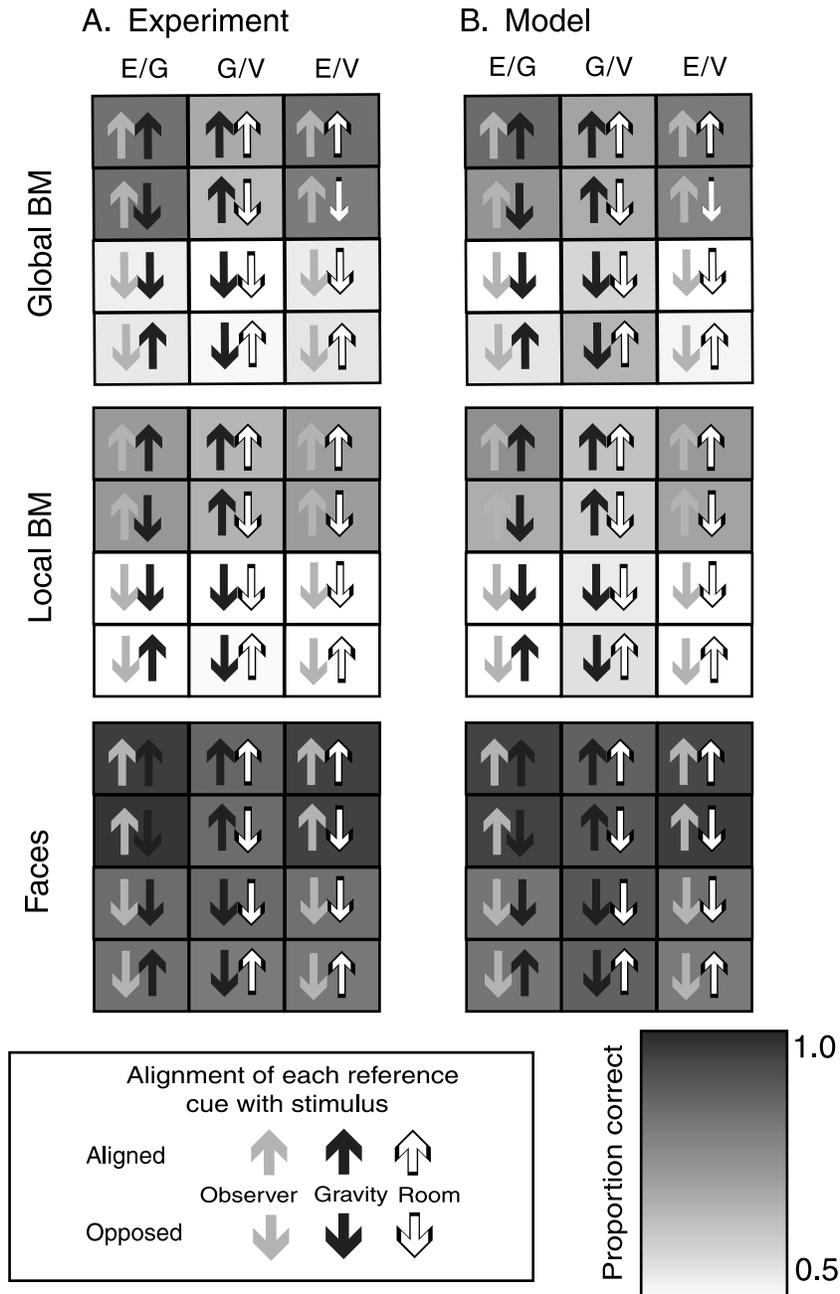


Figure 6. (A) The experimental data are illustrated by the depth of shading for each condition. Cells with no shading represent lowest performance rates (50%) and shading gets darker as performance approaches 100%. (B) The output from the model plotted in the same format. Each column of the matrices corresponds to comparisons between the egocentric and gravitational reference frames (E/G), gravitational and visual environmental reference frames (G/V), and egocentric and visual environmental reference frames (E/V), respectively.

percentages in Table 1. The weightings quantify the role of gravity for each task. For the local and global biological motion tasks, gravity contributes 21–25% of the effect but for faces the contribution is around 2%. The weightings allow us to predict performance for any combination of orientations of visual cues, the body, and gravity.

The predictions for each condition are illustrated in Figure 6B. The variations in performance with reference frame alignment is well matched by the model for all three stimulus types (global BM: $r^2 = 0.20$; local BM: $r^2 = 0.26$; faces: $r^2 = 0.37$).

The role of gravity for biological motion perception

Our data suggest that biological motion is predominantly egocentric-coded. For the retrieval of global structure of an articulated figure, there is no reason to expect involvement of additional allocentric reference frames. However, processing of local motion as a cue to the facing direction of a walker has been shown to be connected to expectations that the visual system has about the dynamics of the body in response to gravitational acceleration (Chang & Troje, 2009a; Shipley, 2003; Troje & Westhoff, 2006).

Of particular interest then is the question of whether such “knowledge” about the effects of gravity is implemented in terms of a heuristic that assumes gravity is aligned with retinal coordinates, or if direct measures of the direction of gravity (taken either by the vestibular system or kinesthetic sensors) feed into the visual system. Our data suggest that there is in fact a small but significant contribution from direct measurements of gravity not only for local motion processing but also for the retrieval of global structure from biological motion.

The contribution of gravity for the perception of biological motion is congruent with the findings by Lopez et al. (2009) that judgments of stability of tilted human body postures also show a small influence of a gravitational reference frame. Direct involvement of the vestibular system for the perception of visual motion, in particular for motion influenced by gravity, has been

	Egocentric	Gravitational	Visual environmental
Global BM	69%	25%	6%
Local BM	71%	21%	8%
Faces	89%	2%	9%

Table 1. Relative weightings of the three reference systems derived from the linear models fit to the biological motion discrimination and face recognition data. These are expressed as a percentage of the total contribution from all three reference frames tested.

demonstrated previously using fMRI (Indovina et al., 2005). In Indovina et al.’s study, motion that was coherent with gravity activated the vestibular network, including areas such as the insular cortex, temporoparietal junction, premotor and supplementary motor areas, middle cingulate cortex, postcentral gyrus, putamen, and the posterior thalamus, in addition to several visual motion areas. In contrast, motion that was inconsistent with gravity elicited responses mainly in the visual areas. These results suggest that the constraints of gravity have been internalized and represented in the brain not just in terms of prior probabilities about the relations between egocentric, visual environmental, and gravitational orientations but also by the integration of direct measurements of the direction of gravity.

Conclusion

Inversion effects for the perception of faces and biological motion have been described pervasively in the literature. The inversion of a stimulus, however, can be described relative to a variety of reference frames including egocentric components (e.g., retina, head, body), gravity, and the visual environment—all of which are usually aligned. Here, we teased apart the roles of these reference frames for the perception of faces and biological motion by systematically misaligning them. We showed that both biological motion and faces are largely coded in an egocentric frame of reference. The relative weightings of the gravitational and visual environmental frames of reference, however, seem to depend on the relevance of the stimulus class to these reference frames. Unlike faces, dynamic events occurring on Earth are affected by gravitational force. Non-visual information about body orientation (reflecting this gravitational component) is integrated into the neural coding of these stimuli.

Acknowledgments

We thank Richard Dyde for his assistance in data collection. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada to NFT and LRH and support from the Canadian Institute for Advanced Research to NFT.

Commercial relationships: none.

Corresponding author: Nikolaus F. Troje.

Email: troje@queensu.ca.

Address: Department of Psychology, Queen’s University, Kingston, Ontario K7L 3N6, Canada.

References

- Bertenthal, B. I., & Pinto, J. (1994). Global processing of biological motions. *Psychological Science*, *5*, 221–225.
- Bingham, G. P., Schmidt, R. C., & Rosenblum, L. D. (1995). Dynamics and the orientation of kinematic forms in visual event recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1473–1493. [PubMed]
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436. [PubMed]
- Chang, D. H. F., & Troje, N. F. (2009a). Acceleration carries the local inversion effect in biological motion perception. *Journal of Vision*, *9*(1):19, 1–17, <http://www.journalofvision.org/content/9/1/19>, doi:10.1167/9.1.19. [PubMed] [Article]
- Chang, D. H. F., & Troje, N. F. (2009b). Characterizing global and local mechanisms in biological motion perception. *Journal of Vision*, *9*(5):8, 1–10, <http://www.journalofvision.org/content/9/5/8>, doi:10.1167/9.5.8. [PubMed] [Article]
- Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception*, *29*, 159–170. [PubMed]
- Howard, I. P. (1982). *Human visual orientation*. New York: Wiley.
- Indovina, I., Maffei, V., Bosco, G., Zago, M., Macaluso, E., & Lacquaniti, F. (2005). Representation of visual gravitational motion in the human vestibular cortex. *Science*, *308*, 416–419. [PubMed]
- 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://www.journalofvision.org/content/3/4/1>, doi:10.1167/3.4.1. [PubMed] [Article]
- Kohler, I. (1940). *Dynamics in psychology*. New York: Liveright.
- Kushiro, K., Taga, G., & Watanabe, H. (2007). Frame of reference for visual perception in young infants during change of body position. *Experimental Brain Research*, *183*, 523–529. [PubMed]
- Lobmaier, J. S., & Mast, F. W. (2007). The Thatcher illusion: Rotating the viewer instead of the picture. *Perception*, *36*, 537–546. [PubMed]
- Lopez, C., Bachofner, C., Mercier, M., & Blanke, O. (2009). Gravity and observer's body orientation influence the visual perception of human body postures. *Journal of Vision*, *9*(5):1, 1–14, <http://www.journalofvision.org/content/9/5/1>, doi:10.1167/9.5.1. [PubMed] [Article]
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. [PubMed]
- Pittenger, J. B. (1985). Estimation of pendulum length from information in motion. *Perception*, *14*, 247–256. [PubMed]
- Rock, I. (1973). *Orientation and form*. New York: Academic Press.
- Rock, I. (1988). On Thompson's inverted-face phenomenon. *Perception*, *17*, 815–817. [PubMed]
- Runeson, S., & Frykholm, G. (1981). Visual perception of lifted weight. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 733–740. [PubMed]
- Shipley, T. F. (2003). The effect of object and event orientation on perception of biological motion. *Psychological Science*, *14*, 377–380. [PubMed]
- Stappers, P. J., & Waller, P. E. (1993). Using the free fall of objects under gravity for visual depth estimation. *Bulletin of the Psychonomic Society*, *31*, 125–127.
- Sumi, S. (1984). Upside-down presentation of the Johansson moving light-spot pattern. *Perception*, *13*, 283–286. [PubMed]
- Thompson, P. (1980). Margaret Thatcher: A new illusion. *Perception*, *9*, 483–484. [PubMed]
- Troje, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision*, *2*(5):2, 371–387, <http://www.journalofvision.org/content/2/5/2>, doi:10.1167/2.5.2. [PubMed] [Article]
- Troje, N. F. (2003). Reference frames for orientation anisotropies in face recognition and biological-motion perception. *Perception*, *32*, 201–210. [PubMed]
- Troje, N. F., & Bühlhoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research*, *36*, 1761–1771. [PubMed]
- Troje, N. F., & Westhoff, C. (2006). The inversion effect in biological motion perception: Evidence for a “life detector?” *Current Biology*, *16*, 821–824. [PubMed]
- Valentine, T. (1988). Upside-down faces: A review of the effect of inversion on face recognition. *British Journal of Psychology*, *79*, 471–491. [PubMed]