What constitutes an efficient reference frame for vision?

Duje Tadin, Joseph S. Lappin, Randolph Blake and Emily D. Grossman

Vanderbilt Vision Research Center, 301 Wilson Hall, Vanderbilt University, 111 21st Avenue South, Nashville, Tennessee 37203, USA
Correspondence should be addressed to D.T. (duje.tadin@vanderbilt.edu)

Vision requires a reference frame. To what extent does this reference frame depend on the structure of the visual input, rather than just on retinal landmarks? This question is particularly relevant to the perception of dynamic scenes, when keeping track of external motion relative to the retina is difficult. We tested human subjects' ability to discriminate the motion and temporal coherence of changing elements that were embedded in global patterns and whose perceptual organization was manipulated in a way that caused only minor changes to the retinal image. Coherence discriminations were always better when local elements were perceived to be organized as a global moving form than when they were perceived to be unorganized, individually moving entities. Our results indicate that perceived form influences the neural representation of its component features, and from this, we propose a new method for studying perceptual organization.

According to the laws of physics, the position and motion of an object can only be defined relative to some reference frame. Neural representations of visual position and motion must abide by the same principles as physical motion, but what is the nature of the reference frame in which the visual system attains efficient representation of position and motion? The nervous system receives visual information in retinocentric coordinates; then this information is transformed into head-centered coordinates for stable perception during eye movements and into a body-centered reference frame to link perception and action[1–3]. Depending on the reference frame, neural representations of motion can be more or less accurate (veridical with respect to the physical world) and more or less efficient (computationally complex). Analogously, planetary motions are both more accurate and more efficient when represented in a heliocentric, rather than a geocentric, reference frame.

Retinal coordinates, along with eye-position correction, are often assumed to be the primary reference frame for neural representations of position and motion. The spatial layout of photoreceptors in the retina is replicated throughout the anatomical hierarchy of visual areas as retinotopically organized maps[4]. This retinotopic organization preserves the original retinal coordinates, which could serve as the reference frame for encoding the motion and position of objects. In the human visual system, however, the motion of an object on the retina does not necessarily imply that the object itself is moving, because our eyes are also usually moving. An accurate retinally based representation would require precise and continuously updated extra-retinal compensation for changes in eye and head position. Vision is known to exploit information in extra-retinal reference signals to compensate for displacements of the retinal image[5–8], but the accuracy of this compensation is debatable[9]. Even if the extra-retinal compensation for changes in eye position is precise, the motions of visual forms and their component features often remain complex; this is a problem for both eye-centered and head-centered representations.

The visual world is dynamic. Spatially separate moving features, such as the arms and legs of an animal, often belong to the same visual form and may even have different trajectories. Individual spots on a leopard's skin will have diverse motion trajectories, which may be very different from the motion of the global form (the whole leopard). The motion of each spot in the retinal reference frame is rather complex and amounts to the vector sum of the observer's motion (eye and body), leopard motion and spot motion. In contrast, if the motion of a spot is visually represented relative to the leopard's form, its motion becomes simpler. Such representation may result in a more accurate perception of visual relationships among local moving features (spots), and culminate in a more efficient perception of the global moving form (leopard).

The motion and position of a feature may be encoded and represented in relation to other stationary or moving objects[10–12], thereby simplifying that feature's visual representation. By definition, this reference frame is non-retinal; it is dynamic and must be continuously updated by new visual input. To what extent do visual reference frames depend on the structure of visual input, not just local retinal coordinates? Can vision bypass or supplement the computational difficulties associated with a retinal reference frame by encoding position and motion relative to perceived forms? This question is especially relevant for perceiving dynamic scenes and for actively exploring one's environment—when the reference frames defined by visual input move relative to the retina.

To determine whether visual information is encoded more efficiently when form information is available, we manipulated the recognizability of motion-defined form while minimizing changes in the retinal image. We modified two well-known protocols: biological motion ('biomotion')[13] and a translating pentagon seen through an aperture mask (adapted from ref. 14). These were chosen because simple manipulations—up/down inversion of biomotion and masking of pentagon apertures—
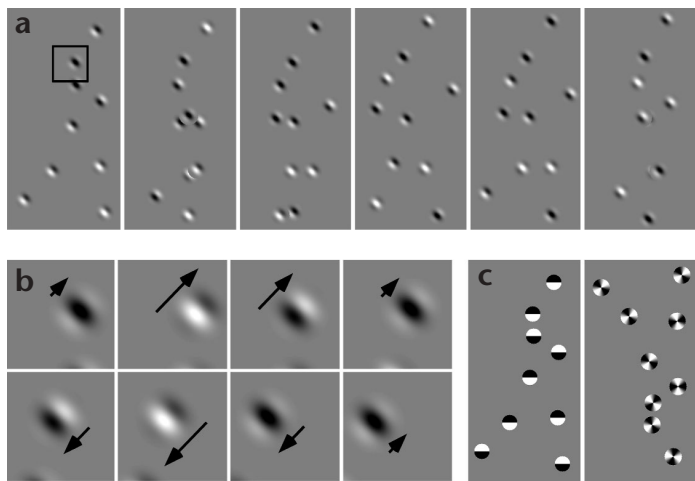
Fig. 1. Point-light walker animations. (**a**) Six frames illustrating the 60-frame point-light walker (PLW) animation defined by Gabor patches. Animation duration was ~1.4 s. Each frame in isolation appears as a random pattern of Gabor patches, but when the animation is set into motion, human form is readily perceived. Sequentially shifting the gaze from frame to frame may give a weak impression of biological motion. In actual experiments, however, observers did not visually pursue the PLW, but fixated at the fixation cross in the center of the screen. (**b**) Eight frames showing a full cycle of 2-Hz oscillatory motion of the grating within the Gabor patch that defines the shoulder of the PLW. The first frame corresponds to the outlined region in panel (**a**). Arrows indicate motion direction and speed of the grating. Note that the position of the entire Gabor patch changes from frame to frame. The magnitude of this position change depends on the Gabor location, with the 'wrist Gabors' undergoing the largest position changes. (**c**) First frames from biological motion animations defined by counterphasing black/white disks (left) and rotationally oscillating windmills (right, illustrating the inverted condition).

transform these perceptually structured moving forms into conglomerations of unorganized elements.

Human form in motion is readily perceived from point-light animations composed of only ~12 points on major joints of the body[13]. If the point-light animation is inverted, human form is no longer perceived although the animation retains its original spatiotemporal structure[15,16]. In masked pentagon displays, the sides of a pentagon are seen through five apertures that occlude its vertices, while the pentagon is translating along a circular path. As a consequence of the aperture problem[17,18], the motions of individual line segments are inherently ambiguous, but observers integrate the motion signals across line segments and a rigidly translating form is readily perceived[14]. If the apertures are invisible, observers perceive only the motions of individual line segments with no apparent global structure.

Both point-light walker (PLW) and masked pentagon (MP) animations were modified for this study. PLW animations were amended by placing small Gabor patches on the major joints of a human walker (**Fig. 1a**). Observers discriminated motion coherence of Gabor patches that oscillated (**Fig. 1b**) either coherently in phase or with some phase difference. These Gabor patches were placed on either upright or inverted PLWs. In analogous experiments, observers judged the coherence of counterphasing black/white disks and rotationally oscillating windmills (**Fig. 1c**). In MP animations, the pentagon was presented behind five apertures and translated along a circular path. Each pentagon side was defined by an oscillating grating oriented parallel to the side (**Fig. 2**). Observers discriminated motion coherence of the five gratings. In separate conditions, the pentagon apertures were either visible (coherent form perceived, **Fig. 2a**) or invisible (no form perceived, **Fig. 2b**). In all experiments, global form (or its absence) was, in principle, irrelevant for the coherence discrimination task.

We manipulated the salience of motion-defined form in PLW and MP animations while making only minimal changes relative to the retinal reference frame. Crucially, context was always irrelevant for performing the tasks. If relative positions and motions of individual elements in perceptually structured displays (upright PLW and MP with visible apertures) are more efficiently encoded because of the availability of form information, perception of spatiotemporal relationships among individual elements should be facilitated. We found that across all displays and tasks, coherence discriminations were more

accurate when the stimulus was perceptually structured, defining a global moving form.

RESULTS
Biological motion

In a series of conditions, we estimated observers' ability to discriminate perceptual coherence of various dynamic elements that defined either upright or inverted PLW animations (Methods and **Fig. 1**). For all tasks and all observers, perceptual coherence judgments were more accurate (thresholds 61% lower) when the local features (Gabors, disks or windmills) defined an upright PLW (**Fig. 3a–e**). The critical difference between the two conditions was that a well-organized global form is perceived in the upright, but not the inverted, PLW condition[15,16].
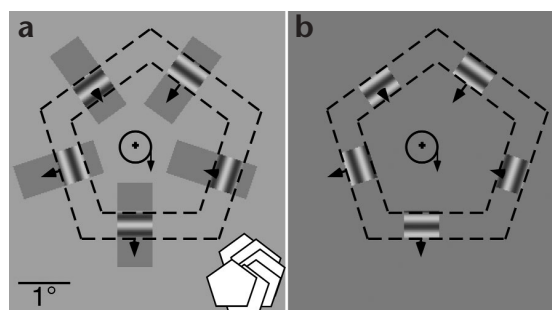

Fig. 2. Translating pentagon animations. (**a**) A translating pentagon was presented behind five apertures (dashed outline is for illustration only). The pentagon translated clockwise along the circular path (as illustrated by the schematic in the bottom right corner). The circle inside the pentagon represents the path taken. The arrow on the circle marks the current position along the path and the direction of translation. The motion of the pentagon results in the back and forth motion of the line segments within apertures. Arrows mark the locations where the line segments would shift as the pentagon moves from the 3 o'clock to the 6 o'clock position. Note that direction and speed vary among different line segments, as depicted by the variable lengths of the arrows. When the apertures were visible (as shown), observers perceived a translating pentagon shape. Independent of the pentagon translation, five gratings oscillated either coherently or incoherently within the limits of pentagon sides. (**b**) Same display except that the luminance of the aperture mask is the same as the background, rendering the apertures invisible. In this condition, observers saw only back and forth motion of the line segments, with no global form information.

2

Observers made perceptual coherence judgments about changing local features that were moving along complex trajectories produced by the PLW animations. (Imagine comparing several similar objects while your friend is juggling them.) In the inverted PLW condition, feature trajectories appear globally unstructured. In upright PLW conditions, the motions of local features are a part of the recognizable global form moving across the screen. If these motions are represented relative to the moving human form, trajectories of local features appear related to each other in a perceptually meaningful way. Evidently, the presence of global moving form facilitates performance by providing a reference frame in which perceptual coherence judgments are easier.

Inversion preserves relative, but not absolute, position and motion of local elements comprising PLW animations (**Fig. 1c**). To verify that our results were indeed due to differences in perceived form, we carried out another version of the initial experiment: observers discriminated motion coherence of oscillating Gabor patches embedded in stationary patterns that were selected from PLW animations. On each trial, Gabor patch positions were assigned on the basis of a single frame that was randomly selected from upright or inverted PLW animation sequences (for example, the third frame from **Fig. 1a**). In these displays, the differences in absolute positions of local elements comprising upright and inverted PLWs are retained, but there is no difference in perceived form—both upright and inverted displays look like randomly positioned oscillating Gabor patches. Other methods were the same as before. Differences in absolute position did not affect performance (**Fig. 3f**).

We set up a control experiment to address the possibility that the presence of biological form may increase attentional engagement[19,20] because upright PLW animations are arguably more interesting than inverted PLWs. This could conceivably result in better performance. In the control study, observers simply detected the rotation of windmills embedded in either upright or inverted PLW animations. Windmills were either stationary or oscillated at 2.4 Hz. The threshold amplitude of oscillation was estimated for inverted PLW and upright PLW conditions. Other methods were the same as before. Unlike the previous tasks, this detection task did not require any perceptual comparisons among multiple features. Scrutiny of just one feature (for example, a windmill placed on the hip of a PLW) would suffice. Therefore, the reference frame provided by the upright PLW was less important, but any attentional benefits of biological motion should have remained. There was no difference, however, between upright and inverted PLW conditions (**Fig. 4**).

Biological motion was used in this study to manipulate the salience of perceived form while minimizing changes in the retinal image. Next, we used very different visual displays to investigate whether this result generalizes to other motion-defined forms.
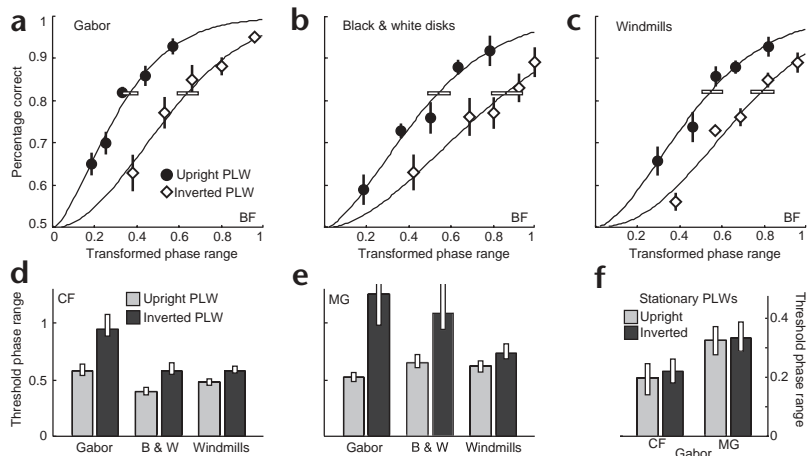

Fig. 3. Results from PLW experiments. (**a**–**c**) Psychometric functions for upright and inverted PLW conditions for observer BF—oscillating Gabor patches (**a**), counterphasing black/white disks (**b**), and rotationally oscillating windmills (**c**). Vertical error bars show s.e.m. for each data point. Horizontal error bars (placed on the 82% point on the psychometric functions) show 95% confidence intervals around threshold estimates. (**d**, **e**) Phase range thresholds (82% correct) for two other observers. Thresholds larger than 1 indicate that observers' accuracy was below the 82% criterion at the maximum possible phase range (360°). Error bars show 95% confidence intervals around threshold estimates. (**f**) Phase range thresholds for oscillating Gabor patches embedded in stationary PLW displays for two observers. Note the reduced range of the *y*-axis, indicating that the thresholds were significantly lower when the Gabor patches were embedded in a stationary pattern. Error bars follow the same convention as (**d**) and (**e**).

Masked pentagon

Results with PLWs were extended by estimating observers' ability to perceive motion coherence of five oscillating gratings, each located at a side of a translating pentagon. The pentagon was presented behind either visible or invisible apertures (Methods and **Fig. 2**). In similar displays (typically a diamond), the visibility of apertures determines whether or not a rigid form is perceived[14]. We used a pentagon shape instead of a diamond because the opposing parallel sides of a diamond are always pairwise-rigid. When viewed through apertures, the sides of a translating pentagon (and any other geometric shape without parallel sides) never move rigidly, thereby potentially enhancing the 'perceived disorganization' when apertures are invisible.

Motion coherence judgments were much better when the apertures were visible and a rigid global form was perceived (**Fig. 5a** and **b**). At the phase range where performance in the condition with visible apertures was almost perfect, performance with invisible apertures was near chance. This result corresponded with a perceptual shift: when the apertures were visible, observers reported seeing a rigid form translating behind the aperture mask, and when the apertures were invisible, they saw disorganized motions of five line segments moving back and forth.

The differences in magnitude of the observed effects in PLW (**Fig. 3a**–**e**) and MP (**Fig. 5a** and **b**) experiments may be due, in part, to the differences in the extent to which perceptual organization was manipulated in each experiment. In MP animations, the contrast was between a rigid moving object (pentagon) and five disorganized line segments moving non-rigidly. In PLW animations, the contrast was between a non-rigid but perceptually well-organized human walker and an equally non-rigid but perceptually disorganized inverted walker. The inverted PLW was not
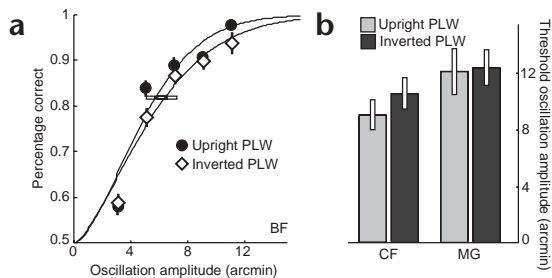
Fig. 4. Results from the motion detection task. (**a**) Psychometric functions for upright and inverted PLW conditions for observer BF. (**b**) Oscillation amplitude thresholds (82% correct) for two other observers. Vertical and horizontal error bars follow the same convention as **Fig. 3**.

perceived as a recognizable form, but it was far from being completely random. Some of its local components are pairwise-rigid (for example, ankle and knee elements), providing some structure. One observer described inverted PLWs as four gravity-defying pendulums. The presence of some structure in inverted PLWs may be responsible for the larger effect seen with MP animations.

Although we attempted to minimize changes in the retinal reference frame for these experiments, retinal images for the two conditions were not identical. Sheer presence of the large aperture mask in one condition might conceivably facilitate performance. To address this possibility, we repeated the experiment with displays in which the pentagon was stationary (the radius of translation was zero). In these displays, conditions with and without visible apertures do not differ in how structured or disorganized they appeared. The presence of the aperture mask did not affect performance (**Fig. 5c** and **d**), and thus cannot explain the results that we attributed to the visibility of form.

Discussion

We developed novel stimuli to manipulate the salience of the form-defined frame of reference while minimizing changes of the retinal image. Our results strongly indicate that when a perceptually organized reference frame is available, the relationships among moving features are represented more accurately. The specific mechanisms underlying this process are not yet evident, but, intuitively, the availability of form may allow for more efficient encoding of relative position and motion. Perceptually organized form provides a well-structured reference frame which may promote a perceptually meaningful representation of the incoming visual input. This is analogous to elaboration effects in human memory: material is better remembered when it is encoded in a meaningful context[21].

The present finding is consistent with other psychophysical results showing that motion perception often depends on the visual context[22–28]. For example, the motion aftereffect is markedly reduced if the motion is presented alone, in the absence of any external reference frame[22], and enhanced if the background is moving in the opposite direction from the adapted motion[26]. The presence of form that might serve as a reference frame in such experiments is confounded, however, by substantial changes in the retinal image (for example, the presence of a conspicuous background versus a blank background). The unique advantage of our approach is that it minimizes this confound and allows us
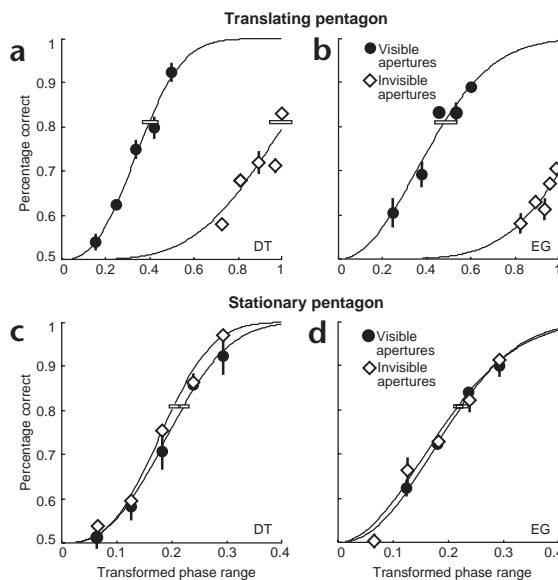
to draw conclusions about the effects of form alone on visual motion representations.

We used two classes of displays whose perceptual organization is well-studied and easily manipulated. In practice, perceptual organization of moving patterns is a difficult problem. Our finding that motion coherence thresholds are reduced when the visual context is a perceptually well-organized moving pattern (such as upright PLW) provides a general method for assessing the 'strength' of perceptual organization. Specifically, if an experimental manipulation of the moving pattern results in reduction of the coherence thresholds, it may be possible to conclude that perceptual organization of the moving pattern has improved. The observation that larger effects were measured with MP animations whose perceptual organization differed more profoundly between conditions suggests that this method is sensitive to the graded disruption of perceptual organization.

We found that the availability of motion-defined form boosts performance in low-level visual tasks. This requires that form information must be available at or before the neural stage(s) where motion coherence and temporal coherence are processed. Evidence indicates that middle temporal visual area (MT or V5) is the neural locus of motion coherence perception[29,30], whereas perception of temporal coherence is generally thought to be an earlier step in visual processing[31]. The present results imply that the neural representation in such early visual areas is influenced by the availability of form. Results from recent psychophysical[32–34] and physiological[35] studies are consistent with an early influence of form in vision. Our results introduce a potentially important function of this early influence: to provide a frame of reference for more accurate representation of the visual input.

Collectively, our results indicate that form has an early influence in visual processing, resulting in a more accurate representation of its component features. Our finding does not

Fig. 5. Results from MP experiments. Psychometric functions for visible and invisible aperture conditions in a motion coherence task for observers DT and EG in the translating (**a, b**) and stationary (**c, d**) pentagon experiments. Note the reduced range of the *x*-axis in (**c**) and (**d**), indicating that when the pentagon was stationary, motion coherence thresholds were significantly reduced. Vertical and horizontal error bars follow the same convention as **Fig. 3**.

imply that the visual system ignores the retinal reference frame. Indeed, retinal representation with extra-retinal compensation is crucial for estimation of heading and visually guided movement[6]. Evidently, when a form-defined reference frame is available, vision is capable of exploiting the additional benefits it provides. These benefits may be greatest while viewing dynamic scenes (as in our displays) or during active exploration of the environment. In such cases, keeping track of motions and positions relative to the retina is difficult, and potential gains from representing at least some of the features relative to visual forms are the greatest. Similar ideas of multiple representations have been advanced to describe how the nervous system transforms sensory input into a representation used by the motor system[1,36,37].

The two classes of displays that we used here are remarkable examples of form-from-motion. Motion, however, is just one of several ways in which form can be defined by the visual system. Luminance, color, texture and stereo cues also contribute significantly to our perception of form[38]. We have focused on moving forms because they are more difficult to represent within the retinal reference frame. Representing stationary objects within retinal coordinates is computationally simpler (particularly in absence of eye movements), and thus potential benefits of form-defined reference frames may be less important. Indeed, our control experiments with stationary patterns show highest performance, with thresholds about half of those for a translating pentagon seen through visible apertures (compare **Fig. 5a** and **b** with **Fig. 5c** and **d**). It is altogether possible that the present results will generalize to other types of form cues, but we speculate that the effects may be smaller than those obtained with motion-defined form.

Human vision has evolved into a flexible neural system that makes use of diverse sources of information. One of these sources is the structure of the visual stimulus itself[10], which may be exploited by the nervous system to obtain an accurate and efficient representation of the visual environment. The results presented here show that the visual system takes advantage of the structure of its input to more efficiently represent relative positions and motions.

METHODS

Stimulus patterns were created with the Psychophysics Toolbox[39,40] on an Apple Macintosh G4 computer (Cuppertino, California). Patterns were displayed on a linearized monitor at 85 Hz. Viewing was binocular and conducted in photopic ambient illumination (4.8 cd/m²). Background luminance was 60.5 cd/m².

Thresholds were estimated by the method of constant stimuli. A session comprised 200 trials, with five conveniently chosen stimulus levels. Three or four sessions were run for each condition and 82% thresholds were estimated by fitting a Weibull distribution[41] to the data. Confidence intervals (95%) around threshold estimates were determined using a bootstrap procedure[42,43]. Trials were self-paced, and feedback was provided for correct responses. Three naïve, paid and well-practiced observers participated in the PLW experiments. Two authors were observers for the MP experiments. All experiments complied with Vanderbilt University Institutional Review Board procedures, and all observers gave informed consent.

Biological motion. Biological motion displays were created by placing ten Gabor patches (σ = 5 arcmin; 4 cycles/°, 92% contrast) on the major joints of a human point-light walker (**Fig. 1**). Each Gabor patch consisted of a moving sine grating windowed by a stationary Gaussian envelope. All Gabor patches had the same orientation, which was randomly selected on each trial. Gabor patches (gratings therein) oscillated sinusoidally (2 Hz, amplitude 180°, starting spatial phase ran-

domized), either coherently or with some phase difference. The observer's task was to report if the Gabor patches oscillated coherently or incoherently. The task difficulty was controlled by adjusting the range of possible phases from which—in incoherent trials—the oscillation phase for each Gabor was randomly selected. When the phase range is sufficiently narrow, incoherently oscillating Gabor patches appear to move coherently. Before thresholds could be estimated, phase range values were transformed to correct for the fact that the average phase difference between Gabor patches increases non-linearly for phase ranges between 180° and 360° (see **Supplementary Methods** and **Supplementary Fig. 1** online).

As Gabor patches were placed on the joints of a PLW, the position of each Gabor patch changed every 70 ms (~14 Hz). This change of position was independent of the 2-Hz Gabor patch oscillation (sampled at 42.5 Hz) and irrelevant for the task. Each trial lasted 1.4 s, during which the PLW walked for 2° (~1.4°/s). The PLW started at 1° eccentricity on either side of fixation and 'walked' through the fovea. At this speed, the PLW (~3.8° tall) appeared to walk naturally. Trials of inverted PLWs were intermixed with upright PLW trials. In inverted PLW trials, observers generally perceived disorganized motion of independent features without any global structure, consistent with previous observations[15,16]. During the analysis, the trials with upright and inverted PLWs were separated and a psychometric function was estimated for each condition.

In an analogous experiment, black/white disks (radius 8 arcmin) were placed on the joints of inverted and upright PLWs (**Fig. 1c**). On the first frame, the top and bottom halves of each disk were either black and white or white and black, randomly assigned. The polarity of the disks was reversed at 2 Hz, either coherently or incoherently. In coherent trials, all disks switched polarity synchronously (with the first switch occurring within the first 500 ms), whereas in incoherent trials, the switching occurred within a certain phase range. The observers' task was to make judgments about the temporal coherence[31,44] of counterphasing black/white disks.

In the third experiment, observers discriminated the perceptual coherence of rotationally oscillating windmills (radial gratings) embedded in PLW animations (**Fig. 1c**). Windmills (radius 9 arcmin, 92% contrast) rotated sinusoidally (2.4 Hz, amplitude 180°, starting spatial phase randomized) either coherently or incoherently with oscillation phases randomly selected from a range of phase lags. The rotation direction of each windmill was randomly assigned every trial, so windmills practically never rotated in the same direction. In coherent trials, windmills switched direction synchronously, but also had identical angular speed at any point in time (but not identical velocity, as oscillation directions were random).

Masked pentagon. In MP displays (**Fig. 2**), an outline of an equilateral pentagon (side length 140 arcmin, side width 40 arcmin) was presented behind five rectangular apertures. The pentagon translated along a circular path (radius 16 arcmin, 2.4 rev/s) for 470 ms. Apertures were fixed in location and placed so that pentagon vertices were always occluded. The foreground of the aperture mask had a luminance that was either identical to the background, rendering the apertures invisible (**Fig. 2b**), or 11.5 cd/m² lighter than the background, resulting in visible apertures (**Fig. 2a**).

Each pentagon side was defined by a sine grating (3.5 cycles/°, 46% contrast) oriented parallel to the side. Within the limits of the pentagon borders, gratings oscillated sinusoidally either coherently or incoherently with some phase difference (2.1Hz, amplitude 180°, starting spatial phase randomized). These oscillations were independent of the line segment motion resulting from pentagon translation. As before, task difficulty was controlled by adjusting the phase range from which oscillation phases were selected in incoherent trials. In a single trial, incoherent and coherent displays were presented in separate temporal intervals and observers identified the interval in which the five gratings oscillated coherently (temporal 2AFC task with an interstimulus interval of 470 ms). Trials with visible and invisible apertures were intermixed, and after the experiment, a psychometric function was estimated for each condition separately.

1. Boussaoud, D. & Bremmer, F. Gaze effects in the cerebral cortex: reference frames for space coding and action. *Exp. Brain Res.* **128**, 170–180 (1999).
2. Buneo, C. A., Jarvis, M. R., Batista, A. P. & Andersen, R. A. Direct visuomotor transformations for reaching. *Nature* **416**, 632–636 (2002).
3. Soechting, J. F. & Flanders, M. Moving in three-dimensional space: frames of reference, vectors, and coordinate systems. *Annu. Rev. Neurosci.* **15**, 167–191 (1992).
4. Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).
5. Bradley, D. C., Maxwell, M., Andersen, R. A., Banks, M. S. & Shenoy, K. V. Mechanisms of heading perception in primate visual cortex. *Science* **273**, 1544–1547 (1996).
6. Crowell, J. A., Banks, M. S., Shenoy, K. V. & Andersen, R. A. Visual self-motion perception during head turns. *Nat. Neurosci.* **1**, 732–737 (1998).
7. Haarmeier, T., Thier, P., Repnow, M. & Petersen, D. False perception of motion in a patient who cannot compensate for eye movements. *Nature* **389**, 849–852 (1997).
8. Haarmeier, T., Bunjes, F., Lindner, A., Berret, E. & Thier, P. Optimizing visual motion perception during eye movements. *Neuron* **32**, 527–535 (2001).
9. Turano, K. A. & Heidenreich, S. M. Eye movements affect the perceived speed of visual motion. *Vision Res.* **39**, 1177–1187 (1999).
10. Gibson, J. J. *The Perception of the Visual World* (Houghton Mifflin, Boston, 1950).
11. Johansson, G. in *Perceiving Events and Objects* (eds. Jaansson, G., Bergstrom, S. S. & Epstein, W.) 29–122 (Lawrence Erlbaum, Hillsdale, New Jersey, 1994).
12. Wade, N. J. & Swanston, M. T. A general model for the perception of space and motion. *Perception* **25**, 187–194 (1996).
13. Johansson, G. Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* **14**, 201–211 (1973).
14. Lorenceau, J. & Shiffrar, M. The influence of terminators on motion integration across space. *Vision Res.* **32**, 263–273 (1992).
15. Fox, R. & McDaniel, C. The perception of biological motion by human infants. *Science* **218**, 486–487 (1982).
16. Pavlova, M. & Sokolov, A. Orientation specificity in biological motion perception. *Percept. Psychophys.* **62**, 889–899 (2000).
17. Marr, D. & Ullman, S. Directional selectivity and its use in early visual processing. *Proc. R. Soc. Lond. B Biol. Sci.* **211**, 150–180 (1981).
18. Morgan, M. J., Findlay, J. M. & Watt, R. J. Aperture viewing: a review and a synthesis. *Q. J. Exp. Psychol.* **34A**, 211–233 (1982).
19. Cavanagh, P., Labianca, A. T. & Thornton, I. M. Attention-based visual routines: sprites. *Cognition* **80**, 47–60 (2001).
20. Vaina, L. M., Solomon, J., Chowdhury, S., Sinha, P. & Belliveau, J. W. Functional neuroanatomy of biological motion perception in humans. *Proc. Natl. Acad. Sci. USA* **98**, 11656–11661 (2001).
21. Craik, F. I. M. & Tulving, E. Depth of processing and the retention of words in episodic memory. *J. Exp. Psychol. Gen.* **104**, 268–294 (1975).
22. Day, R. H. & Strelow, E. Reduction or disappearance of visual after effect of movement in the absence of patterned surround. *Nature* **230**, 55–56 (1971).
23. Lappin, J. S. & Craft, W. D. Foundations of spatial vision: from retinal images to perceived shapes. *Psychol. Rev.* **107**, 6–38 (2000).
24. Lappin, J. S., Donnelly, M. P. & Kojima, H. Coherence of early motion signals. *Vision Res.* **41**, 1631–1644 (2001).
25. Legge, G. E. & Campbell, F. W. Displacement detection in human vision. *Vision Res.* **21**, 205–213 (1981).
26. Murakami, I. & Shimojo, S. Modulation of motion aftereffect by surround motion and its dependence on stimulus size and eccentricity. *Vision Res.* **35**, 1835–1844 (1995).
27. Nakayama, K. & Tyler, C. W. Relative motion induced between stationary lines. *Vision Res.* **18**, 1663–1668 (1978).
28. Nawrot, M. & Sekuler, R. Assimilation and contrast in motion perception: explorations in cooperativity. *Vision Res.* **30**, 1439–1451 (1990).
29. Newsome, W. T., Britten, K. H. & Movshon, J. A. Neuronal correlates of a perceptual decision. *Nature* **341**, 52–54 (1989).
30. Shadlen, M. N., Britten, K. H., Newsome, W. T. & Movshon, J. A. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J. Neurosci.* **16**, 1486–1510 (1996).
31. Lee, S. -H. & Blake, R. Visual form created solely from temporal structure. *Science* **284**, 1165–1168 (1999).
32. Croner, L. J. & Albright, T. D. Image segmentation enhances discrimination of motion in visual noise. *Vision Res.* **37**, 1415–1427 (1997).
33. Lorenceau, J. & Alais, S. Form constraints in motion binding. *Nat. Neurosci.* **4**, 745–751 (2001).
34. Verghese, P. & Stone, L. S. Perceived visual speed constrained by image segmentation. *Nature* **381**, 161–163 (1996).
35. Croner, L. J. & Albright, T. D. Segmentation by color influences responses of motion-sensitive neurons in the cortical middle temporal visual area. *J. Neurosci.* **19**, 3935–3951 (1999).
36. Battaglia-Mayer, A. *et al.* Early coding of reaching in the parieto-occipital cortex. *J. Neurophysiol.* **83**, 2374–2391 (2000).
37. Snyder, L. H. Coordinate transformations for eye and arm movements in the brain. *Curr. Opin. Neurobiol.* **10**, 747–754 (2000).
38. Regan, D. M. *Human Perception of Objects* (Sinauer, Sunderland, Massachusetts, 2000).
39. Brainard, D. H. The psychophysics toolbox. *Spat. Vis.* **10**, 443–446 (1997).
40. Pelli, D. G. The Video Toolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).
41. Quick, R. F. Jr. A vector-magnitude model of contrast detection. *Kybernetik* **16**, 65–67 (1974).
42. Wichmann, F. A. & Hill, N. J. The psychometric function: I. Fitting, sampling, and goodness of fit. *Percept. Psychophys.* **63**, 1293–1313 (2001).
43. Wichmann, F. A. & Hill, N. J. The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Percept. Psychophys.* **63**, 1314–1329 (2001).
44. Blake, R. & Yang, Y. Spatial and temporal coherence in perceptual binding. *Proc. Natl. Acad. Sci. USA* **94**, 7115–7119 (1997).