

Chapter 33

Biological and Body Motion Perception

Martin A. Giese

Section for Computational Sensomotrics, Dept. of Cognitive Neurology,
Hertie Institute for Clinical Brain Research & Center for Integrative Neuroscience
University Clinic Tübingen
Otfried-Müller-Str. 25, 72076 Tübingen, Germany
Email: martin.giese@uni-tuebingen.de

Abstract:

The recognition of biological motion and actions is a central visual function that requires the spatio-temporal integration of complex visual patterns. This function requires advanced mechanisms of pattern recognition and the organization of Gestalt patterns in space-time. The investigation of body motion perception has a long tradition in psychology. However, clear ideas about underlying neural and computational mechanisms, and their neural implementations, have only emerged recently. This research shows that body motion perception requires an integration of multiple visual processes, including Gestalt-like pattern formation, learning-based recognition, and the interaction between bottom-up and top-down processing

A huge variety of empirical studies have been collected that treat different aspects of the perception of biological and body motion, ranging from psychophysical questions, the processing of social signals, over ecological and developmental aspects, to clinical implications. Due to space limitations, this chapter focuses primarily on aspects related to pattern formation and the organization of Gestalt for dynamic patterns.

Many topics in body motion perception, which cannot be covered in this chapter due to space limitations, are treated in many excellent review articles and books. This includes the original work by Gunnar Johansson (review: Jansson et al., 1994), the neural basis of body and facial motion processing (Puce & Perrett, 2003; Alison et al. 2000; O'Toole et al. 2002), computational principles (Giese & Poggio, 2003), imaging results (Blakemore & Decety, 2001; Puce & Perrett, 2003), and its relationship to emotion processing (de Gelder, 2006).

Another important topic that cannot be adequately treated in this review due to space limitations is the relationship between body motion perception and motor representations. Several recent books treat exhaustively different aspects of biological and body motion perception, which could not be included in this review (e.g. Knoblich et al. 2006; Johnson & Shiffrar et al. 2013; Rizzolatti & Sinigaglia, 2008).

Historical background

While already Aristotle had written about the principles of movements of animals, the systematic scientific investigation of body motion perception started back two centuries ago with the works of Eadweard Muybridge (1887) and Etienne-Jules Marey (1894) who studied body motion, applying the technique of sequential photography. While classical Gestalt psychologists had treated the organization of complex motion patterns not so extensively, the systematic study of biological and body motion was initiated by the Swedish psychologist Gunnar Johansson in the 1970s. He was originally interested in studying Gestalt laws of motion organization, and for him body motion was an example of a complex motion pattern with relevance for everyday life (Jansson et al. 1994). His work on biological motion grew out of studies on the organization of much simpler motion patterns during his PhD thesis (Johansson, 1950), aiming at the development of a general ‘theory of event perception’.

Already classical Gestalt psychologists had described pattern organization phenomena for simple motion patterns. This includes the classical law of ‘common fate’ (Wertheimer, 1923), work on motion grouping (Ternus, 1926) and on ‘induced motion’ by Duncker (1929) (see Figure 1A), and studies by Metzger (1937) on the ‘Prägnanz’ in motion perception (see Herzog & Ogmen, this volume). In addition, some more recent work by Albert Michotte (1946/1963) addressed the interpretation of simple motion displays in terms of the perception of ‘causality’.

Johansson tried to study systematically Gestalt grouping principles in simple motion displays that consisted of small numbers of moving dots, where he varied systematically their geometrical and temporal parameters. A variety of his observations are in-line with modern theories about the estimation of optic flow from spatio-temporal image data, such as the tendency to group dots with similar motion vectors in the image plane, or a tendency to favor correspondences in terms of slow motion.

In addition, Johansson made the important additional discovery that he formalized in his theory of *vector analysis*: Often even simple motion patterns are perceptually organized in terms of interpretations that impose a hierarchy of spatial frames of reference, instead of a simple perceptual representation that reflects just the physical structure of the motion. Some example stimuli that illustrate this phenomenon are shown in Figure 1B-D. The physical motion of the stimulus is decomposed into components that describe, sometimes non-rigid deformations within the grouped structure (e.g. a contracting bar), and a second motion component that describes the motion of the whole grouped structure within the external frame of reference (e.g. the movement of the whole bar). The key point is that the perceptual interpretation provides a description in terms of *relative motion* that is described

within frames of reference, which partially result from the grouping process itself. This can be interpreted as a form of vectorial decomposition of the motion, e.g. in a component that describes the motion of a whole group of dots, and an additive second vectorial component that describes the relative motion between the individual dots within the groups. It seems obvious that the principle might be extendable for more complex displays, e.g. consisting of multiple non-rigid parts that move against each other. The human body is an example for such a more complex system, and this motivated originally the interest of Johansson in these types of stimuli.

The analysis of such hierarchical patterns of relative motion is an interesting theoretical problem, and has motivated theoretical work in psychology that tried to account for the organization of such patterns by an application of coding theory and the principle of minimum description length (Restle, 1979). The underlying idea is to characterize different possible encodings of the motion patterns by the required number of describing parameters (such as amplitude, phase, and frequency for sinusoidal oscillation). Encodings in terms of hierarchies of relative motions are often more compact, i.e. require less describing parameters than the direct encoding of the physical movements. In computer vision the minimum description length principle has been successfully applied, e.g., for motion segmentation (Shi & Yu, 1998) and the compression of motion patterns in videos (e.g. Nicolas et al. 1997). However, general models that decompose complex motion patterns in terms of hierarchies of relative motion, in the way envisioned by Johansson, remain to be developed.

Psychophysical investigation of biological and body motion perception

One of the most famous discoveries by Gunnar Johansson was that body motion can be recognized from motion patterns that present only moving dots at the positions of the joints of moving humans, in absence of any information about the body surface (Johansson, 1973). He generated these stimuli by fixing light bulbs or reflecting tapes on the major joints of his participants and filming them in the dark (Figure 2), a technique that was originally developed by Murray. (Today such stimuli are typically generated by motion capture (data bases see, e.g., Vanrie & Verfaillie, 2004; Ma et al. 2006). Johansson's unexpected observation was that observers were able to recognize body motion easily from such strongly impoverished stimuli, even if they were presented only for a very short time (such as 200 ms) (Johansson, 1976). Static patterns of this type, however, could not be easily interpreted by the observers.

Phenomenological studies

Subsequent early research on body motion perception verified that different *categories of movements* could be recognized from point-light stimuli, such as walking, running, or

dancing (e.g. Johansson, 1973; Dittrich, 1993). Further studies showed that humans also can recognize animals, such as or dogs from such point-light stimuli (e.g. Bellefeuille & Faubert, 1998; Jokisch & Troje, 2003). Many early experiments tried to characterize the capability to derive *subtle information* from such motion cues, such a gender (Barclay et al. 1978; Cutting et al. 1978; Pollick et al. 2005), gaits of familiar people or friends (e.g. Beardsworth & Buckner, 1981; Cutting & Kozlowski, 1977), age (Montpare & Zebrowitz-McAthur, 1988), or emotions (e.g. Dittrich et al. 1996; Walk & Homan, 1984; Atkinson et al. 2004; Roether et al. 2009). Also, it has been shown that observers can derive physical properties, such as the weights of lifted objects from such point-light stimuli (e.g. Runeson & Frykholm, 1981). In the context of these early studies, also the first mathematical descriptions for critical features, e.g. for gender perception, and simplified mathematical models for gait trajectories, suitable for the synthesis of point-light pattern by computer graphics (Cutting et al. 1978) have been developed. In addition, minimum coding theory was extended to gait patterns (Cutting, 1981).

Already starting to investigate the underlying critical processes, another stream of experiments investigated the *robustness* of the perception of body motion from point-light stimuli, introducing specific manipulations of Johansson's original stimuli. This includes the masking of point-light stimuli by moving dot masks, generated from randomly positioned moving dots from point-light stimuli ('scrambled walker noise') (Bertenthal & Pinto, 1994; Cutting et al. 1978). Other studies tried to degrade the local motion information by introducing temporal delays between the stimulus frames (Thornton et al. 1998), variations of contrast polarity and spatial-frequency information, or by changing the relative phase of the dots or their disparity information (Ahlström et al. 1997). The depth information in binocularly presented point-light stimuli could be strongly degraded without the observers even noticing this manipulation (Bülthoff et al. 1998). This observation seems incompatible with mechanisms of biological motion recognition that rely on a veridical reconstruction of depth. However, more recent studies show that depth has an important influence and can disambiguate bistable point-light stimuli whose orientation in space cannot be uniquely derived from two-dimensional information (Vanrie et al. 2004; Jackson & Blake, 2010). Other studies tried to degrade point-light stimuli by randomizing the positions of the dots on the body (Cutting, 1981) and by limiting the life time of individual dots (e.g. Neri et al. 1998; Beintema & Lappe, 2002). Another interesting manipulation looking specifically for the organization of biological motion patterns in terms of spatial units were studies that randomized the position of individual parts of the body, leaving their internal motion invariant (showing e.g. all limbs, vs. only the ipsi- or contralateral limbs) (Pinto & Shiffrar, 1999; Neri et al. 2009).

Finally, another set of studies used the rotation of point-light walkers in the image plane ('*inversion*') in order to study frames of reference in which the underlying perceptual processing happens. Like for the perception of faces, rotations in the image plane strongly

degrades the perception of body motion from point-light stimuli (e.g. Sumi, 1984; Pavlova & Sokolov, 1990). The orientation dependence seems to be tied to an egocentric rather than to the external frame of reference (e.g. Troje, 2003). Also the ‘Thatcher illusion’ (that is the difficulty to recognize inverted face parts in faces that are presented upside down) has been generalized to biological motion patterns (Mirenzi & Hiris, 2011). In line with this, a recent study has shown that the features of the local dots (e.g., color) are less accessible for consciousness when they are embedded in an upright than in an inverted biological motion walker (Poljac et al., 2012). These results strongly suggest that the perceptual processing of biological motion might be critically dependent on templates that are tied to the visual frame of reference, rather than on a generic process that reconstructs three-dimensional shape from motion.

Continuous perceptual spaces of motion

The relevance of learned templates in the processing of biological and body motion is also supported by the observation of gradual generalization between different similar body motion patterns. A hallmark of such generalization is an encoding in terms of topologically well-defined *perceptual spaces*.

In computer graphics, for a long time blending techniques have been applied for the generation of novel movements with intermediate style properties. An example are ‘gait designers’ for the generation of gender-specific walking or of body movements with different emotional styles (e.g. Unuma et al. 1995; Wiley & Hahn, 1997; Rose et al. 1998). Psychologists have used similar techniques to generate *style spaces* of body motion in order to study of the perception and categorization of movements (Pollick et al. 2001; Hill & Pollick, 2000; Giese & Lappe, 2002; Troje, 2002). As for faces, it has been shown that body movements can be made particularly expressive and discriminable by extrapolation in such style spaces (‘caricature effect’). As for object recognition (Bülthoff & Edelman, 1992), the categorization of motion patterns seems to be characterized by smooth generalization fields (Giese & Lappe, 2002). In addition, the metric properties of the underlying perceptual space can be recovered by applying multi-dimensional scaling to similarity judgments for body motion patterns, finding that its metric closely resembles to the one defined by distance measures in space-time between the trajectories. This implies a ‘veridical’ encoding of the physical properties of body motions in such perceptual spaces (Giese et al., 2008).

Neural representations of continuous topological pattern spaces give rise to high-level after effects. This has been first shown for static pictures of faces (Leopold et al. 2001). Adaptation with an ‘anti-face’ (a face located opposite to the original face, relative to the average face, in face space) results in an after-effect: The average face is then briefly perceived as the original face immediately after the adaptation phase. Similar after-effects have been observed for biological motion: If for example observers are exposed to a

female walker for several seconds, they perceive a gender-neutral morph temporarily as male walk (Jordan et al. 2006; Troje et al. 2006). It has been shown that such after-effects are not simply a reflection of low-level form or motion after-effects, and must be based on higher representations of body motion. Recent studies have started to investigate how form and motion representations contribute to such high-level after-effects (Theusner et al. 2011).

From critical features to ‘life detectors’

A substantial amount of research in the field of biological motion perception has been searching for the visual features that might be critical for the perception of body motion. At the same time, this work has isolated different levels of the analysis of body motion.

A prominent example of this is work about the relevance of *form vs. motion features*. While some studies, in-line with Johansson’s original inspiration, have provided evidence for a critical role of motion features (e.g. Cutting et al., 1988; Mather et al. 1992; Thornton et al. 1998; Neri et al. 1998; Casile & Giese, 2005), others have strongly emphasized the role of form information (e.g. Beintema & Lappe, 2002; Hiris, 2007). It is critical in this context to define precisely what ‘form’ and ‘motion information’ means, and what exactly is understood by ‘recognizing body motion’. Figure 3A-C tries to illustrate different cues in the processing of body motion. One type of form-based information is the global configuration of the human body. Information about body shape seems at least critical for recognizing moving bodies in clutter, such as in randomly moving background elements (e.g. Lu, 2010). However, such global configurations can be specified based on local form features (panel A), as well as on local motion features (panel B) (specifying complexly structured optic flow patterns). It is thus a logical error to confuse the relevance of the body configuration with an exclusive relevance of shape information. An alternative to the processing of the global configural shape, which is sufficient to solve certain tasks (e.g. to detect body parts, or whether a walker is going right or left,) is the use of local features, or even individual dot trajectories (panel C). Such tasks can be solved without necessarily perceiving a whole human body, e.g. by detection of asymmetry in the motion.

The fact that it is easy to recognize walking or running from static pictures of stick figures shows that form information is relevant for the processing of body motion (Todd, 1983). In addition, it seems obvious that humans can learn to recognize point-light configurations, just as any other shape, after sufficient training (Reid et al. 2009). Computational work has tried to identify critical features for body motion perception, which generalize spontaneously from full-body figures to point-light stimuli, applying principle components analysis to motion and form features. It turns out that such generalization is easier to achieve for motion than for form features (Casile & Giese, 2005). In addition, the opponent motion of the hand and the feet seems to be a critical feature for the recognition of biological motion (Casile & Giese, 2005; Chang & Troje, 2009). Trying to oppose the

potential relevance of local motion cues, Beintema and Lappe (2002) have demonstrated that point-light walkers can be recognized from stimuli where the dot positions are randomized on the skeleton in every frame. This manipulation degrades the local motion information, but does not eliminate some of the critical motion features (Casile & Giese, 2005).

While Lappe and colleagues hypothesized that local motion processing is completely irrelevant for biological motion processing, unless the moving figure has to be segmented from a (stationary) background (Lange & Lappe, 2006), studies comparing the relevance of form and motion cues sometimes found a primary relevance of form and sometimes of motion cues (e.g. Lu & Liu, 2006; Hiris et al., 2007; Thurman & Grossman, 2008). Instead of denying the relevance of individual cues, more recent work has rather studied how the cues are integrated. A recent set of studies tried to develop reverse correlation techniques in order to identify critical features that drive the categorization of biological motion patterns (Lu & Liu, 2006; Thurman & Grossman, 2008; Thurman et al. 2010). These studies found evidence for a relevance of both types of features, consistent with the hypothesis that the nervous system fuses different informative cues during the processing of body motion (instead of dumping classes of informative cues). Further evidence suggests that it is dependent on the task which cue is more effective (Thirkettle et al. 2009). In the same direction points also a recent study that suggests the existence of separate high-level after-effects that are dependent on form or motion cues (Theusner et al. 2011).

A further stream of research about features in the recognition of body motion has been initiated by the observation that the walking direction of point-light walkers can even be derived from scrambled walkers, for which the configural information about the body shape has been destroyed. In addition, the recognition of walking direction from these stimuli is worse if these stimulus patterns are rotated upside down, implying an inversion effect (Troje & Westhoff, 2006). The fact that the walking direction can be recognized without the configural information in a forced-choice task is due to the fact that in particular the foot movement trajectory of walking is highly asymmetrical (Figure 3C). (This fact is analogous to the observation that it is easy to detect the facing direction of side views of faces from only the direction in which the nose points, see Figure 3D.) The recognition of walking direction from such individual dot trajectories is consistent with motion template detectors that are defined in a retinal frame of reference. It is unclear in how far such detectors are learned or partially innate. Some researchers have interpreted the above observation as evidence for a special-purpose mechanism for the detection of the asymmetric foot trajectories, which has been termed '*life detector*'. Since a similar inversion effect was observed for the tendency of newly hatched chicks to align their bodies with point-light patterns (Vallortigara & Regolin, 2006), it has also been hypothesized that this special purpose mechanism is evolutionary old, and potentially

universal through a lot of species. (See also Koenderink's chapter on Gestalts as ecological templates, this volume.) The concept of the 'life detector' has initiated a number of follow-up studies, investigating the processing of biological motion information in absence of configural cues. For example, the perceived temporal duration of biological motion and scrambled biological motion is prolonged compared to similar non-biological stimuli (Wang & Jiang, 2012).

A further general approach for the characterization of signals that are specific for biological movements, and which can be processed even in absence of configural cues, has been motivated by work in motor control on the differential invariants of body movements. An example for such an invariant is the *two-thirds power law* that links the speed and the curvature of the endpoint trajectories of arm and finger movements, and which holds even for trajectories in locomotion. Psychophysical and imaging work shows that trajectories compatible with this law are perceived as smoother (Viviani et al. 1989; Bidet-Ildei et al. 2006), and activate brain structures involved in body motion processing more strongly than dot trajectories that are incompatible with this invariant (Dayan et al. 2010; Casile et al. 2011).

Bottom-up vs. top-down processing

Since a long time, there has been a discussion in the field of body motion perception about possible contributions of bottom-up vs. top-down mechanisms. 'Bottom-up mechanisms' are typically understood as processes that derive representations of complex pattern by combination of simpler image features, e.g. using hierarchical representations. 'Top-down processing' is typically understood as a class of mechanisms that either tries to match some higher representation, e.g. of a moving body to the stimulus sequence, or which actively searches and groups components of body motion stimuli in the stimulus sequence. Typically, it is assumed that these processes require attention.

Initial studies investigated the influence of attention on biological motion processing, demonstrating that biological motion perception tolerates longer inter-stimulus intervals (ISIs) than would be expected from first-order local motion processing (Thornton et al. 1998) and that that processing of biological motion requires attention in dual task and visual search paradigms (Figure 4A) (Cavanagh et al. 2001; Thornton et al. 2002). Consistent with this idea, patients with parietal lesions are impaired in visual search tasks with biological motion stimuli (Battelli et al. 2003). In a more recent study that demonstrates top-down interactions in the processing of biological motion by (Hunt & Halper, 2008) the dots of a normal point-light walker were replaced by complex objects (cf. Figure 4B). This manipulation interfered strongly with the processing of body motion, potentially because attentional resources have to be shared between object and body motion processing.

A substantial attentional modulation of the brain activity related to biological motion processing is also suggested by fMRI and ERP studies (Safford et al. 2010). More detailed psychophysical studies showed that in particular performance variations due to changes of flanker congruency and Stroop-related attention tasks correlated with performance in biological motion processing, while this was not the case for other attention tasks (Chandrasekaran et al. 2010). However, even unattended, not task-relevant walkers are processed automatically in a flanker paradigm and influence the processing of the attended stimulus (Thornton & Vuong, 2004). This illustrates that the control by attention is not complete, and that even in tasks that require top-down control, bottom-up processes act in parallel.

Further experiments show that the processing of body motion interacts with other perceptual processes, and the processing of the scene. For example, the perception of the direction of ambiguous background motion (suggesting a floor or wall) is biased by the perceived locomotion direction of walkers (cf. Figure 4C) (Fujimoto, 2003; Fujimoto & Yagi, 2008). Also, Gestalt grouping principles interact with the perceptual organization of biological motion displays. This was, for example, demonstrated by replacing the dots of point-light walkers by oriented Gabor patches that support or disfavor the correct grouping into limbs (Poljac et al. 2011).

Relevance of learning

Several studies that the perception of body motion and other complex motion patterns is dependent on learning. It is a classical result that observers can learn to recognize individuals from their body movements (e.g., Hill & Pollick, 2000; Kozlowski & Cutting, 1977; Troje, Westhoff, & Lavrov, 2005). The discrimination of biological from scrambled patterns can be successfully trained, where this training induces corresponding changes of the BOLD activity in relevant areas (Grossman et al. 2004). Several studies have compared the learning of biological and similar non-biological motion patterns, finding substantial learning effects, for both stimulus classes (Hiris et al. 2005; Jastorff et al. 2006). It seems critical for the learning process that the learned patterns are related to an underlying skeleton. Beyond this, the learning seems to be very fast, requiring less than 30 repetitions, and it is associated with BOLD activity changes along the whole visual pathway (Jastorff et al. 2009). Finally, the learning of the visual discrimination of body motion patterns has been studied extensively in the context of different application domains. For example, experience seems to improve body motion recognition of identity and emotional expression in dance (e.g. Sevdalis & Keller, 2011), or the efficiency of the prediction of dangerous events in surveillance videos (e.g. Troscianko et al. 2004).

Related to the role of learning in body motion recognition is the question about the extent in which this capability is innate, and how this capability has changed in the course of evolution. This question is on the one hand addressed by many developmental studies,

showing that the capability to discriminate point-light from scrambled stimuli emerges very early in child development (e.g. Fox & McDaniel, 1982; Bertenthal, 1993). Space does not permit to provide a more detailed review of this interesting literature. In addition, a variety of studies has investigated biological motion perception in other species, such as cats, pigeons, or macaques (e.g. Blake, 1993; Dittrich et al. 1998). While many species can discriminate intact point-light from scrambled stimuli more detailed investigations suggest that even macaques might not perceive point-light stimuli in the same way as humans do and require extensive training until they can recognize these patterns correctly (Vangeneugden et al. 2010). This makes it crucial to carefully dissociate the relevant computational levels of the processing of body motion in such experiments with other species, before drawing far-reaching conclusions about potential evolutionary aspects.

Neural mechanisms

Electrophysiological studies

Substantial insights have been gained about neural mechanisms that are involved in the processing of body motion. In particular, the imaging literature on action processing is vast, and a review would by far exceed the scope of this chapter. In the following only a few key results from monkey physiology and functional imaging can be highlighted that are particularly relevant for aspects of visual pattern organization. In addition, it will not be possible to discuss the relevant literature from neuropsychology and the relationship between body motion perception, brain lesions, and psychiatric disorders, such as autism. More comprehensive discussions can be found in reviews about the neural basis of body motion processing (e.g. Decety & Grezes, 1999; Vaina et al. 2004; Puce & Perrett, 2003; Knoblich et al. 2006; Blake & Shiffrar, 2007; Johnson & Shiffrar, 2013).

Neurons with visual selectivity for body motion and point-light stimuli have been first described in the superior temporal sulcus (STS) by the group of David Perrett (Perrett et al. 1985; Oram et al. 1996). This region contains neurons that respond selectively to human movements and body shapes, and in the monkey likely represents a site of convergence of form and motion information along the visual processing stream. Some neurons in this area show specific responses to combinations of articulatory and translatory body motion, and many of them show selectivity for the temporal order of the stimulus frames (Jellema & Perrett, 2003; Barraclough et al. 2009). The responses of many of these neurons are specific for certain stimulus views, and such view dependence has been observed even at very high levels of the processing pathway, e.g. in mirror neurons in premotor cortex (Caggiano et al. 2011). An extensive study of the neural encoding of body motion in the STS has been realized by Vangeneugden et al. (2009) using a stimulus set that was generated by motion morphing, and defining a triangular configuration in the morphing space. Applying multi-dimensional scaling to the responses of populations of STS neurons, corresponding metric configurations in the ‘neural space’ were recovered from the cell activities that closely resembled these configurations in the physical space (consistent with

a veridical neural encoding of the physical space). In addition, this study reports ‘motion neurons’, especially in the upper bank and fundus of the STS, which respond to individual and small groups of dots in point-light stimuli, even in absence of global shape information. Conversely, the lower bank contains many ‘shape neurons’ that are specifically selective for the global shape of the body. Recent studies also applied neural decoding approaches using classifiers to responses of populations of STS neurons for stick figure stimuli, as well as for densely textured avatars, showing that such stimuli can be decoded from such population responses (Singer & Sheinberg, 2010; Vangeneugden et al. 2011). Another literature in the field of electrophysiology that is highly relevant for body motion processing is related to the ‘mirror neuron system’, and shows that neurons in parietal and premotor cortex also are strongly activated by the observation of body motion. Space limitation do not permit here to give a thorough review of this aspect, and the reader is referred to reviews and books that treat specifically this aspect (e.g. Rizzolatti et al. 2001; Rizzolatti & Craighero, 2004; Rizzolatti & Sinigaglia, 2008).

Imaging studies

Meanwhile there exists a vast imaging literature on the perception of body motion, and we can highlight only a very small number of aspects related to the mechanisms of pattern formation. Further details can be found in the reviews mentioned at the beginning of this chapter.

Early positron emission spectroscopy (PET) and fMRI studies found evidence for the involvement of a network of areas, including the posterior STS, in the processing of point-light biological motion (Bonda et al. 1996; Vaina et al. 2001; Grossman & Blake, 2002). The relevant network includes also human MT, parts of the lateral occipital complex (LOC), and the cerebellum. Also an inversion effect could be demonstrated for the activity in the STS (Grossman & Blake 2001). Subsequent studies tried to dissociate activation components related to the action vs. human shape (Peuskens et al. 2005), where specifically the right pSTS seems to respond selectively to the human motion. The human STS can also be robustly activated by full-body motion patterns (e.g. Pelphrey. 2003), and several studies have investigated body motion-induced activation patterns using natural stimuli such as movies (e.g. Hasson et al. 2004; Bartels & Zeki 2004), even being able to decode semantic categories from action videos (Huth et al. 2012). TMS stimulation in the STS reduces the sensitivity to biological motion stimuli (Grossman et al. 2005).

Substantial work has been dedicated to study of body-selective areas in the inferotemporal cortex and their involvement in the processing of body motion. One such area is the extrastriate human body area (EBA) (Peelen & Downing 2007), which is selectively activated by static body shapes and responds also strongly to body motion. Another relevant area is the fusiform body area (FBA), which is very close to the fusiform face area (FFA) (Peelen & Downing, 2005). Both areas have been interpreted as specifically

processing the form aspects of body motion. Recent studies, controlling for structure as well as motion cues, suggests that EBA and FBA might represent an essential stage of body motion processing that links the body information with the action (Jastorff & Orban, 2009). Very similar imaging results have been obtained by fMRI studies in the monkey cortex, permitting to establish a homology between human and monkey imaging data on body motion perception (e.g. Jastorff et al. 2012).

Again, there exists a vast and continuously growing imaging literature about the involvement of motor and mirror representations in the perceptual processing of body motion. Again we refer to other more specialized reviews (e.g. Buccino et al., 2004; van Overwalle & Baetens, 2009) with respect to this aspect.

Computational and neural models

Motion recognition and tracking have been popular topics in computational and computer vision since the 1990s, and a huge variety of algorithms have been developed in this domain. Only a small number of these approaches is relevant for biological systems. For a recent overview over technical approaches see e.g. Moeslund et al. (2006). We will briefly sketch here some computational approaches that have been developed in the psychological literature on body motion perception, and we will then more thoroughly discuss existing neural models.

Computational models

Early theories of body motion recognition were based on simple invariants that can be derived from the three-dimensional movements of articulated figures (e.g., Hoffman & Flinchbaugh, 1982; Webb & Aggarwal, 1982). For example, for point-light stimuli the distances between dots on the same limb tend to vary less than the distances between dots on different limbs. Alternatively, one can try to derive geometrical constraints for the two-dimensional motion of points that are rigidly connected in the three-dimensional space. Classical work by Marr and Vaina (1982), assumed that the brain might recover the body shape, and track body movements, using parametric body models that are composed from cylindrical shape primitives. Other models have exploited other shape primitives, such as spheres (e.g. O'Rourke & Badler, 1980).

Building on this idea another class of theoretical models has been developed that is presently very influential in cognitive neuroscience. This class of models assumes that the recognition of body movements and actions is based on the *internal simulation of observed motor behaviors*. A tight interaction between body motion recognition and motor control is suggested by many experiments (reviews see e.g. Knoblich et al. 2006; Schütz-Bosbach & Prinz, 2007). For example, a study by Jacobs & Shiffrar (2005) shows that the perception of gait speeds of point-light walkers depends on whether the observers are walking or running during the observation. A direct and highly selective coupling between motor

control and mechanisms for the perception of biological motion is also suggested by a study that used Virtual Reality technology in order to control point-light stimuli by the concurrent movements of the observer (e.g. Christensen et al. 2011). In this case, detection of biological motion was facilitated if the stimulus was spatially and temporally coherent with the ongoing movements of the observer, but impaired if this congruency was destroyed. In addition, a variety of studies demonstrate that motor expertise (independent of visual expertise) influences performance in body motion perception (e.g. Hecht et al. 2001; Casile & Giese, 2006; Calvo-Merino et al. 2006)

The analysis-by-synthesis idea that underlies this class of models goes back to classical *motor theory of speech recognition*, which assumes that perceived speech is mapped onto ‘vocal gestures’ that form the units of the production of speech in the vocal tract (Lieberman et al. 1967). For action recognition this idea has been formulated, for example, by Wolpert and colleagues who suggested that controller models for the execution of body movements might be used also for motion and social recognition (Wolpert et al. 2003). The underlying idea is illustrated in Figure 5A. Their MOSAIC model is based on a mixture of controller experts (forward models) for the execution of different behaviors. Recognition is accomplished by predicting the observed sensory signals using all controller models, and selecting the one that generates the smallest prediction error. Models based on similar ideas have been suggested as account for the function of the ‘mirror neuron system’ in action recognition, and as basis for the learning of movements by imitation (e.g. e.g. Oztup & Arbib, 2002; Erlhagen et al. 2006). In addition, related models have also been formulated exploiting a Bayesian framework (e.g. Kilner et al. 2005).

Many of the discussed analysis-by-synthesis approaches require the reconstruction of motor-relevant sensory variables, such as joint angles, at the input level. The estimation of such variables from monocular image sequences is a very difficult computer vision problem that is partially unsolved. Correspondingly, only few of the discussed models are implemented to a level that would demonstrate their performance on real video data. For the brain it is unclear if and how it solves the underlying reconstruction problem. Alternatively, the visual system might circumvent this difficult computational problem, recognizing body motion by computationally simpler strategies.

Neural models

Another class of models has been inspired by fundamental properties of the architecture of the visual cortex and extends biologically-inspired models for the recognition of stationary shapes (e.g. Riesenhuber & Poggio, 1999) in space-time. Such an architecture, which reproduces broad range of data about body motion recognition from psychophysics, electrophysiology, imaging, and neuropsychology is illustrated in Figure 5B. (See Giese & Poggio (2003), Casile & Giese (2005), Giese (2006), Fleischer et al. (2013) for a detailed description.) Consistent with the anatomy of the visual cortex, the model is organized in terms of two hierarchical neural pathways, modeling the ventral and dorsal processing

streams. The first pathway is specialized for the processing of form information, while the second pathway processes local motion information.

Both pathways consist of hierarchies of neural detectors that mimic properties of cortical neurons, and which converge to a joint representation at a level that corresponds to the STS. The complexity of the extracted features as well as the receptive field sizes of the feature detectors increase along the hierarchy. The model creates *position and scale invariance* along the hierarchy by pooling of the responses of detectors for the same feature over different positions and scales, using a maximum operation (e.g., Riesenhuber & Poggio, 1999). Stimuli can thus be recognized largely independently of their size and positions in the visual field.

The detectors in the *form pathway* mimic properties of shape-selective neurons in the ventral stream (including simple and complex cells in primary visual cortex, V4 neurons, and shape-selective neurons in infero-temporal cortex). The detectors on the highest level of the form pathway ('snapshot neurons') are selective body postures that are characteristic for snapshots from movies showing the relevant body movement. They are modeled by radial basis function (RBF) units, which represent a form of fuzzy shape template (the RBF center defining the template). The *motion pathway* of the model has the same hierarchical architecture, where its input level is formed by local motion energy detectors. This pathway recognizes temporal sequences of complexly-structured optic flow patterns, which are characteristic for body motion.

A central idea of the model is that body motion can be recognized by identifying temporal sequences of features, such as body shapes or optic flow patterns in 'snapshots' from a movie (Giese, 2000). In order to make the neural detectors selective for the temporal order of such sequences, the model assumes the existence of asymmetric lateral connections between the snapshot neurons in the form and motion pathway. The resulting network dynamics suppresses responses to movies for which the stimulus frames appear in the wrong temporal order (Giese & Poggio, 2003).

The model accomplishes recognition first in a *view-specific manner*, within view-specific modules that are trained with different views of the body motion sequence. Only on the highest hierarchy level the information from different view-specific modules is combined by pooling, resulting in *view-independent* motion recognition (cf. Figure 5B).

If such a model is trained with normal full-body motion and tested with point-light walkers the motion pathway spontaneously generalizes to point-light stimuli, while this is not the case for the form pathway. This does not imply that configural information is irrelevant because also the optic flow templates in the motion pathway are dependent on the global body configuration. In addition, this result does not imply that the form pathway cannot

process point-light patterns. If trained with them, the form pathway responds also perfectly to dot patterns (Casile & Giese, 2005), consistent with the fact that trained observers can learn to recognize actions even from static point-light patterns (Reid et al. 2009).

A strongly related model has been proposed by Beintema and Lappe (2006). This model was designed originally in order to account for the processing of a biological motion from stimuli that degrade local motion information by repositioning the dots on the skeleton of a moving point-light figure in every frame (Beintema & Lappe, 2002). This model is very similar to the form pathway of the model by Giese & Poggio (2003), where the major differences are: (i) The model does not contain a motion pathway; (ii) it does contain a mechanism that accounts for position and scale invariance; and (iii) it implicitly assumes that the form template detectors (RBFs) are always perfectly positioned and scaled relative to the stimulus. In presence of static backgrounds this perfect alignment might be accomplished by motion segmentation (Lange & Lappe, 2006), while this approach seems not applicable in presence of motion clutter, e.g. for dynamically masked point-light stimuli. (More extensive discussions of related models can be found in Giese (2006) and Fleischer et al. (2013).)

Meanwhile, much more computationally efficient versions of the Giese-Poggio model have been developed in computer vision, reaching state-of-the-art performance for action detection (e.g. Jhuang et al. 2007; Escobar et al. 2009; Schindler et al. 2008). In addition, the model has been extended for *the recognition of goal-directed actions* (Fleischer et al. 2013). For this purpose, additional modules were integrated that model the properties of neurons in parietal and premotor cortex. One of these modules computes the spatial relationship (relative position and motion) between the moving effector (e.g. the hand) and the goal object. The other module contains neurons (probably in the STS and parietal cortex) that combine the information about the goal object, the effector movement, and the spatial relationship between effector and goal. The model accomplishes recognition of goal-directed hand actions from real videos, at the same time reproducing a whole spectrum of properties of action-selective neurons in the STS, parietal and the premotor cortex. Opposed to the architecture shown in Figure 5A, recognition by this model is accomplished without the explicit reconstruction of three-dimensional structure parameters, such as joint angles, from monocular image sequences. In addition, it has been shown (Fleischer et al. 2012) that the model even accounts for certain forms of causality perception (Michotte, 1946/1963).

Conclusion

This chapter has reviewed some central results and theories about the perception of body motion. Work on this topic in psychology started from the original work of Johansson, who studied body motion as an example of complex and ecologically relevant natural motion, and who was aiming at uncovering and testing Gestalt rules for the perceptual organization

of motion. Since then, this field has made a strong development during which it has absorbed many other approaches outside Gestalt psychology and pattern formation. This includes psychophysical theories of pattern detection, top-down control by attention, learning-based recognition theories, ecological and developmental psychology, and modern approaches in physiology and imaging, including neural decoding by machine learning techniques. The large body of existing work has revealed some neural and computational principles. However, we have no clear picture of the underlying neural and computational processes, and many of existing explanations remain phenomenological, theoretically not rigorously defined, or only loosely tied to experimental data. The main stream of present research is dominated, on the one hand, by pattern recognition approaches, implicitly assuming signal detection or filtering mechanisms, partly combined with ecological ideas. Contrasting with this approach, research in cognitive neuroscience is fascinated by the idea of an analysis by internal simulation of motor behavior, often entirely bypassing the aspects of visual pattern recognition. Both streams go away from Johansson's original idea of uncovering the dynamic processes that control pattern formation in the organization of complex motion patterns. It seems likely that such processes play a central role in the organization of ambiguous stimulus information about body motion, and it seems quite interesting to pick up this old line of research. Modern mathematical approaches in neurodynamics, Bayesian inference, and computational learning, combined with the now available computer power, will provide a methodological basis to re-address these questions. This approach in this direction seems even more promising since the previous work has revealed insights about relevant features and underlying basic processes, laying a basis for the study of active pattern formation in the processing of naturalistic body motion stimuli.

Acknowledgments

I thank M. Angelovska for help with the illustrations and the editing of the references. I thank J. Vangeneugden and an anonymous reviewer for helpful comments. Supported by EU Commission, EC FP7-ICT-248311 AMARSi, F7 7-PEOPLE-2011-ITN: ABC PITN-GA-011-290011, HBP FP7-ICT-2013-FET-F/ 604102; FP7-ICT-2013-10/ 611909 KOROIBOT, Deutsche Forschungsgemeinschaft: DFG GI 305/4-1, DFG GZ: KA 1258/15-1, and German Federal Ministry of Education and Research: BMBF, FKZ: 01GQ1002A.

References

- Ahlström, V., Blake, R., & Ahlström, U. (1997). Perception of biological motion. *Perception*, 26, 1539-1548.
- Allison, T., Puce, A., McCarthy, G. (2000). Social perception from visual cues: role of the STS

- region. *Trends Cogn Sci.* 4,267–278.
- Atkinson, A.P., Dittrich, W.H., Gemmel, A.J., & Young A.W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, 33, 717-746.
- Barclay, C., Cutting, J., & Kozlowski, L. (1978). Temporal and spatial factors in gait perception that influence gender recognition. *Percept. Psychophys.* 23, 145–152.
- Barracough, N.E., Keith, R.H., Xiao, D., Oram, MW, & Perrett, D.I. (2009). Visual adaptation to goal-directed hand actions. *J. Cogn. Neurosci*, 21, 1806–1820.
- Bartels, A., Zeki, S. (2004). Functional brain mapping during free viewing of natural scenes. *Hum. Brain Mapp.* 21:75–85.
- Battelli, L., Cavanagh, P., & Thornton, I.M. (2003). Perception of biological motion in parietal patients. *Neuropsychologia* 41, 1808–1816.
- Beardsworth, T., & Buckner, T. (1981). The ability to recognize oneself from a video recording of one's movements without seeing one's body. *Bulletin of the Psychonomic Society*, 18, 19–22.
- Bellefeuille, A., & Faubert, J. (1998). Independence of contour and biological-motion cues for motion-defined animal shapes. *Perception* 27, 225–35.
- Beintema, J.P., & Lappe M. (2002). Perception of biological motion without local image motion. *Proceedings of the National Academy of Science USA*, 99, 5661–5663.
- Beintema, JA, Georg, K, Lappe, M. (2006). Perception of biological motion from limited lifetime stimuli. *Percept. Psychophys.* 68(4), 613-624.
- Bertenthal, B. I. (1993). Perception of biomechanical motions by infants: Intrinsic image and knowledge-based constraints. In C. Granrud (Ed.), *Carnegie Symposium on Cognition: Visual perception and cognition in infancy* (pp. 175-214). Hillsdale, NJ: Erlbaum.
- Bertenthal, B. I., & Pinto, J. (1994). Global processing of biological motions. *Psychological Science*, 5, 221–225.
- Bidet-Ildes C., Orliaguet J. P., Sokolov A. N., & Pavlova M. (2006). Perception of elliptic biological motion. *Perception*, 35, 1137–1147.
- Blake, R. (1993). Cats perceive biological motion. *Psychological Science* 4, 54-57.
- Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annu Rev Psychol.* 58, 47-73.
- Blakemore, S.J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nat. Rev. Neurosci.* 2, 561– 66.
- Bonda, E., Petrides, M., Ostry, D., Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *J Neurosci.* 16(11), 3737-44.
- Bülthoff, I., Bülthoff, H., Sinha, P. (1998). Top-down influences on stereoscopic depth-perception. *Nat. Neurosci.* 1, 254–257.
- Bülthoff, H.H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences*, 89, 60-64.
- Buccino, G., Binkofski, F., & Riggio, L. (2004). The mirror neuron system and action recognition. *L. Brain Lang.* 89(2):370-376.
- Calvo-Merino, B, Grèzes, J, Glaser, DE, Passingham, RE, Haggard, P.I. 2006 Seeing or doing? Influence of visual and motor familiarity in action observation. *Curr Biol.* 2006 Oct 10;16(19):1905-10.
- Caggiano, V, Fogassi, L, Rizzolatti, G, Pomper, J, Thier, P, Giese, M.A., Casile, A (2011) View-based encoding of actions in mirror neurons of area f5 in macaque premotor cortex. *Curr Biol*

- 21:144–148.
- Casile, A., & Giese, M.A. (2005). Critical features for the recognition of biological motion. *Journal of Vision*, 5, 348-360.
- Casile, A., & Giese M. A. (2006). Non-visual motor learning influences the recognition of biological motion. *Current Biology*, 16(1), 69-74.
- Casile, A., Dayan, E., Caggiano, V., Hendler, T., Flash, T., Giese, M.A. (2011). Neuronal encoding of human kinematic invariants during action observation. *Cereb Cortex* 20(7):1647-1655.
- Cavanagh, P., Labianca, A.T., Thornton, I.M. (2001). Attention-based visual routines: sprites. *Cognition*, 80:47–60.
- Chang, D.H., & Troje, N.F. (2009) Acceleration carries the local inversion effect in biological motion perception. *J Vis* 9(1):19.1-17.
- Chandrasekaran C., Turner L., Bülthoff H. H., & Thornton I. M., (2010). Attentional networks and biological motion, *Psihologija*, 43(1), 5-20.
- Christensen, A., Ilg, W. & Giese, M. A. (2011). Spatiotemporal Tuning of the Facilitation of Biological Motion Perception by Concurrent Motor Execution. *Journal of Neuroscience*, 31(9), 3493-3499.
- Cutting, J. E., & Kozlowski, L. T., (1977) Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society*, 9, 353 -356.
- Cutting, J.E., Proffitt D.R., & Kozlowski L.T. (1978). A biomechanical invariant for gait perception. *Jornal of Experimental Psychology: Human Perception and Performance* 4, 357-372.
- Cutting, J. E. (1981). Coding theory adapted to gait perception. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 71-87.
- Cutting, J.E., Moore, C., Morrison, R. (1988). Masking the motions of human gait. *Percept.Psychophys.* 44, 339–347.
- Dayan, E., Casile, A., Levit-Binnun, N., Giese, M.A., Hendler, & T., Flash, T. (2010). Neural representations of kinematic laws of motion: evidence for action-perception coupling. *Proc Natl Acad Sci U S A*, 104(51), 20582-20587.
- Decety, J., & Grèzes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends Cogn Sci.* 3(5), 172-178.
- de Gelder B. (2006). Towards the neurobiology of emotional body language. *Nat Rev Neurosci.* 7(3), 242-249
- Dittrich, W.H. (1993). Action categories and the perception of biological motion. *Perception* 22,15–22.
- Dittrich, W. H., Troscianko, T., Lea, S. E., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception*, 25, 727–738.
- Dittrich, W.H., Lea, S.E.G., Barrett, J., & Gurr, P.R. (1998). Categorization of natural movements by pigeons: visual concept discrimination and biological motion. *J. Exp. Anal. Behav.* 70:281–99.
- Duncker, K. (1929). Über induzierte Bewegung (Ein Beitrag zur Theorie optisch wahrgenommener Bewegung). *Psychologische Forschung*, 12, 180-259.
- Erlhagen W, Mukovskiy A, Bicho E. (2006). A dynamic model for action understanding and goal-directed imitation. *Brain Res.* 1083(1),174-188.
- Escobar, M.J., Masson, G.S., Vieville, T., Kornprobst, P. (2009.) Action recognition using a bio-

- inspired feedforward spiking network. *Int. J. Comput. Vision* 82, 284–301.
- Fleischer F, Christensen A, Caggiano V, Thier P, Giese MA. (2012). Neural theory for the perception of causal actions. *Psychol Res.* 76(4), 476-493.
- Fleischer, F., Caggiano, V., Thier, P. & Giese, M. A. (2013). Physiologically inspired model for the visual Recognition of transitive hand actions. *Journal of Neuroscience, The Journal of Neuroscience.*, 15(33), 6563-80.
- Fox, R., & Mc Daniel, C. (1982). The perception of biological motion by human infants. *Science.* 218(4571), 486-7.
- Fujimoto, K. (2003). Motion induction from biological motion. *Perception* 32:1273–1277.
- Fujimoto, K., Yagi, A. (2005). Motion illusion in video images of human movement. In: F. Kishino et al. (eds.): *ICEC 2005, LNCS 3711*, Springer-Verlag, Berlin / Heidelberg, 531–534.
- Fujimoto, K., Yagi, A. (2008). Biological motion alters coherent motion perception. *Perception.* 37(12), 1783-1789.
- Giese, M.A., & Lappe, M. (2002). Measurement of generalization fields for the recognition of biological motion. *Vision Res.* 42(15),1847-1858.
- Giese, M.A. (2000). Neural field model for the recognition of biological motion patterns. *Second Proceedings of International ICSC Symposium on Neural Computation (NC 2000)*, 1-12.
- Giese, M.A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nat Rev Neurosci* , 4, 179–192.
- Giese, M.A. (2006). Computational Principles for the Recognition of Biological Movements, Model-based versus feature-based approaches. Knoblich, W., Thornton, I. M., Grossjaen, M., Shiffrar, M. (eds): *The Human Body: Perception From the Inside Out*. Oxford University Press, 323-359.
- Giese, M. A., Thornton, I.M., & Edelman, S. (2008). Metrics of the perception of body movement. *Journal of Vision*, 8(9), 1-18.
- Grossman, E.D., Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vision Res.* 41(10-11), 1475-1482.
- Grossman, E.D., Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron.* 35(6), 1167-1175.
- Grossman ED, Blake R, Kim CY. (2004). Learning to see biological motion: brain activity parallels behavior. *J. Cogn. Neurosci.* 16:1669–79.
- Grossman, E.D., Battelli, L., & Pascual-Leone A. (2005). Repetitive TMS over STSp disrupts perception of biological motion. *Vis. Res.* 45:2847–53.
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, 303,1634–401.
- Hecht, H., Vogt, S., & Prinz, W. (2001). Motor learning enhances perceptual judgment: a case for action-perception transfer. *Psychol Res.* 65(1), 3-14.
- Herzog, M. H., & Öğmen, H. (2014). Apparent motion and reference frames. In J. Wagemans (Ed.), *Oxford Handbook of Perceptual Organization* (in press). Oxford, U.K.: Oxford University Press.
- Hill, H. & Pollick, F.E. (2000). Exaggerating temporal differences enhances recognition of individuals from point light displays. *Psychological Science* Vol.11 (3), 223-228.
- Hiris E, Krebeck A, Edmonds J, Stout A. 2005. What learning to see arbitrary motion tells us about biological motion perception. *J. Exp. Psychol.: Hum. Percept. Perform.* 31:1096–106
- Hiris, E. (2007). Detection of biological and nonbiological motion. *J Vis.* 7(12), 4.1-16.

- Hoffman, D.D., Flinchbaugh, B.E. (1982). The interpretation of biological motion. *Biol Cybern.* 42(3), 195-204.
- Hunt, A.R., & Halper, F. (2008). Disorganizing biological motion. *J Vis.* 8(9),12.1-5.
- Huth, A.G., Nishimoto, S., Vu, A.T., & Gallant, J.L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron.* 76(6), 1210-1224.
- Jackson, S., & Blake, R. (2010) Neural integration of information specifying human structure from form, motion, and depth. *J Neurosci.* 30(3), 838-48.
- Jacobs, A., & Shiffrar, M. (2005). Walking perception by walking observers. *J. Exp. Psychol.: Hum. Percept. Perform.* 31,157–69.
- Jansson, G., Bergström, S.S., Epstein, W., & Johansson, G. (1994). *Perceiving Events and Objects.* Lawrence Erlbaum Associates: Hillsdale, NJ.
- Jastorff, J., Kourtzi, Z., Giese, M.A. (2006). Learning to discriminate complex movements: biological versus artificial trajectories. *J Vis.*,6(8):791-804.
- Jastorff, J., Kourtzi, Z., & Giese, M.A. (2009). Visual learning shapes the processing of complex movement stimuli in the human brain. *J Neurosci.* 29(44),14026-14038.
- Jastorff, J., & Orban, G.A. (2009). Human functional magnetic resonance imaging reveals separation and integration of shape and motion cues in biological motion processing. *J Neurosci.* 29(22), 7315-7329.
- Jastorff, J., Popivanov, I.D., Vogels, R., Vanduffel, W., & Orban, G.A. (2012). Integration of shape and motion cues in biological motion processing in the monkey STS. *Neuroimage.* 60(2), 911-921.
- Jellema, T., & Perrett, D.I. (2003). Perceptual history influences neural responses to face and body postures. *J Cogn Neurosci.* 15(7), 961-71.
- Johansson, G. (1950). *Configurations in event perception: an experimental study*, Diss. Stockholm : Höskolan, Stockholm.
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14, 201–211.
- Johansson, G. (1976). Spatio-temporal differentiation and integration in visual motion perception An experimental and theoretical analysis of calculus-like functions in visual data processing. *Psychological Research*, 38, 379–393.
- Johnson, K. & Shiffrar, M. (2013). *People Watching.* Oxford University Press.
- Jokisch, D., & Troje, N.F. (2003). Biological motion as a cue for the perception of size. *J.Vis.* 3,252–264.
- Jordan H, Fallah M, Stoner GR. (2006) Adaptation of gender derived from biological motion. *Nat Neurosci.* 9(6),738-739.
- Kilner, J., Friston, K.J., & Frith, C.D. (2005). The mirror-neuron system: a Bayesian perspective. *Neuroreport.* 18(6), 619-623.
- Knoblich, G., Thornton, I.M., Grosjean, M., & Shiffrar, M. (2006). *Human Body Perception from the Inside Out.* New York: Oxford Univ. Press.
- Koenderink, J. (2014). Gestalts as ecological templates. In J. Wagemans (Ed.), *Oxford Handbook of Perceptual Organization* (in press). Oxford, U.K.: Oxford University Press.
- Lange, J., & Lappe, M.(2006). A model of biological motion perception from configural form cues. *J Neurosci* 26, 2894–2906.
- Leopold, D.A., O’Toole, A.J., Vetter, T., & Blanz, V. (2001). Proto-type-referenced shape

- encoding revealed by high-level aftereffects. *Nat. Neurosci.* 4, 89–94.
- Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol Rev.* 74(6),431-61.
- Lu, H., & Liu, Z. (2006). Computing dynamic classification images from correlation maps. *J Vis.* 6(4), 475-83.
- Lu, H. (2010). Structural processing in biological motion perception. *J Vis.* 10(12), 1-13.
- Ma, Y., Paterson, H.M., & Pollick, F.E. (2006). A motion-capture library for the study of identity, gender, and emotion perception from biological motion. *Behav. Res. Methods* 38, 134–41.
- Marey, E.J. (1894). *Le Mouvement*, Masson, Paris.
- Marr, D, & Vaina, L. (1982). Representation and recognition of the movements of shapes. *Proc R Soc Lond B Biol Sci.* 214(1197), 501-24.
- Mather, G., Radford, K., & West, S. (1992). Low level visual processing of biological motion. *Proc. R.Soc. Lond. B Biol. Sci.* 249,149–55.
- Metzger, W. (1937). 'Gesetze des Sehens' (1st German edition, 1937; *Laws of Vision*)
- Michotte, A. (1946). *La perception de la causalité*. Publications Universitaires Louvain. (English translation: *The perception of causality*. Methuen, London (1963).)
- Mirenzi, A., & Hiris, E., (2011). The Thatcher effect in biological motion. *Perception* 40(10), 1257 – 1260.
- Moeslund, T.B., Hilton, A., & Kruger, V. (2006). A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding* 104, 90–126.
- Montpare, J. M., Zebrowitz, M., & McArthur, L. (1988). Impressions of people created by age-related qualities of their gaits. *Journal of Personality and Social Psychology*, 55, 547–556.
- Muybridge, E. (1887). *Muybridge's Complete Human and Animal Locomotion*. (All 781 Plates from the 1887 "Animal Locomotion." Volume I. Dover Publications, Inc. 1979.)
- Nicolas, H., Pateux, S., & Le Guen, D. (1997). Minimum description length criterion for region-based video compression, *Image Processing, Proceedings, International Conference* 1, 346-349 .
- Neri, P., Morrone, M.C., & Burr D. (1998). Seeing biological motion. *Nature* 395, 894–96.
- Neri, P. (2009). Wholes and subparts in visual processing of human agency. *Proc Biol Sci.* 276(1658), 861-869.
- Oram, M.W., Perrett, D.I. (1996). Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *J. Neurophysiol.* 76:109–129.
- O'Rourke J. & Badler N. (1980). "Model-based image analysis of human motion using constraint propagation." *IEEE Trans. on Pattern Analysis and Machine Intelligence* 2(6),522-536.
- O'Toole, A.J., Roark, D.A., & Abdi, H. (2002). Recognizing moving faces: a psychological and neural synthesis. *Trends Cogn Sci.* 6 (6), 261-266.
- Oztop, E., & Arbib, M.A.(2002). Schema design and implementation of the grasp-related mirror neuron system. *Biol Cybern.* 87(2), 116-40.
- Pavlova, M, & Sokolov, A. (2000). Orientation specificity in biological motion perception. *Percept Psychophys.*, 62 (5), 889-99.
- Peelen, M.V., & Downing, P.E. (2005). Selectivity for the human body in the fusiform gyrus. *J Neurophysiol.* 93(1):603-8.
- Peelen, M.V., & Downing, P.E.(2007). The neural basis of visual body perception. *Nat Rev Neurosci.* 8(8), 636-48.
- Pelphrey, K.A., Mitchell, T.V., Mc Keown, M.J., Goldstein, J., Allison, T., & Mc Carthy, G.

- (2003). Brainactivity evoked by the perception of human walking: controlling for meaningful coherentmotion. *J. Neurosci.* 23:6819–25.
- Perrett, D.I., Smith, P.A., Mistlin, A.J., Chitty, A.J., Head, A.S., Potter, D.D., Broenni-Mann, R., Milner, A.D., & Jeeves, M.A. (1985). Visual analysis of body movements by neurons in the temporal cortex of the macaque monkey: a preliminary report. *Behav Brain Res* 16, 153–170.
- Peuskens, H., Vanrie, J., Verfaillie, K., & Orban GA. (2005). Specificity of regions processing biologicalmotion. *Eur. J. Neurosci.* 21:2864–75.
- Pinto, J., & Shiffrar, M. (1999). Subconfigurations of the human form in the perception of biologicalmotion displays. *Acta Psychol.* 102,293–318.
- Poljac, E., Verfaillie, K., & Wagemans, J. (2011) Integrating biological motion: the role of grouping in the perception of point-light actions. *PLoS ONE* 6(10), e25867.
- Poljac, E., de-Wit, L., & Wagemans, J. (2012). Perceptual wholes can reduce the conscious accessibility of their parts. *Cognition*, 123, 308-312.
- Pollick, F.E., Paterson, H.M., Bruderlin, A., & Sanford, A.J. (2001). Perceiving affect from arm movement. *Cognition.* 82(2), B51-61.
- Pollick, F.E., Kay, J.W., Heim, K., & Stringer, R. (2005). Gender recognition from point-light walkers. *J. Exp. Psychol.: Hum. Percept. Perform.* 31, 1247–65.
- Puce, A., & Perrett, D., (2003). Electrophysiology and brain imaging of biological motion. *Philos.Trans. R. Soc. Lond. B Biol. Sci.* 358, 435–45.
- Reid, R, Brooks, A, Blair, D, & van der Zwan, R.(2009). Snap! Recognising implicit actions in static point-light displays. *Perception.* 38(4), 613-6.
- Riesenhuber, M, Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat Neurosci.* 12(11):1019-1025.
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2:661–70
- Rizzolatti, G.,& Craighero, L.(2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27,169–192.
- Rizzolatti, G. & Sinigaglia, C. (2008) *Mirrors in the brain: How our minds share actions and emotions.* Oxford University Press, USA.
- Restle, F. (1979) Coding theory of the perception of motion configurations. *Psychol Rev.* 86(1), 1-24.
- Roether, C.L., Omlor, L., Christensen, A. & Giese, M. A. (2009). Critical features for the perception of emotion from gait. *Journal of Vision*, 9(6), 1-32.
- Rose, C., Cohen, M.F., & Bodenheimer, B. (1998). Verbs and adverbs: multidimensional motion interpolation. *Computer Graphics and Applications*, 18(5), 32-40.
- Runeson, S., & Frykholm, G. (1981). Visual perception of lifted weight. *J. Exp. Psychol.: Hum. Percept.Perform.* 7,733–740.
- Safford, A.S., Hussey E.A., Parasuraman, R., & Thompson, J.C. (2010). Object-based attentional modulation of biological motion processing: spatiotemporal dynamics using functional magnetic resonance imaging and electroencephalography. *J Neurosci.*30 (27), 9064-73.
- Schindler, K., Van Gool, L., de Gelder, B. (2008). Recognizing emotions expressed by body pose: a biologically inspired neural model. *Neural Netw.* 21(9),1238-46.
- Schütz-Bosbach, S., & Prinz, W.(2007). Perceptual resonance: action-induced modulation of perception. *Trends Cogn Sci.* 11(8), 349-55.
- Shi, J., Pan, J., & Yu, S. (1998). Joint motion estimation and segmentation based on the MDL

- principle. ICSP '98. Fourth International Conference on Signal Processing, Proceedings, 2(2), 963-967.
- Singer, J.M., Sheinberg, D.L. (2010). Temporal cortex neurons encode articulated actions as slow sequences of articulated poses. *J Neurosci*, 30, 3133–3145.
- Sevdalis, V., & Keller, P.E. (2011). Perceiving performer identity and intended expression intensity in point-light displays of dance. *Psychol Res.* 75(5):423-34.
- Sumi, S. (1984). Upside-down presentation of the Johansson moving light-spot pattern. *Perception* 13, 283–86.
- Theusner, S., de Lussanet, M.H.E., & Lappe, M. (2011). Ad -aptation to biological motion leads to a motion and a form after effect. *Atten Percept Psychophys*, 73(6), 1843–1855.
- Thornton, I. M., Pinto J., & Shiffrar, M. (1998). The visual perception of human locomotion. *Cognitive Neuropsychology*, 15, 535–552.
- Thornton, I.M., Rensink, R.A., & Shiffrar, M.(2002) Active versus passive processing of biological motion. *Perception*, 31(7), 837-53.
- Thornton, I.M., & Vuong, Q.C. (2004.) Incidental processing of biological motion. *Curr Biol.* 14(12), 1084-9.
- Thurman, S.M., & Grossman, E.D. (2008). Temporal "Bubbles" reveal key features for point-light biological motion perception. *J Vis.* 8(3), 28.1-11.
- Thurman, S.M., Giese, M.A., & Grossman, E.D. (2010). Perceptual and computational analysis of critical features for biological motion. *J Vis* 10,1–15.
- Ternus, J. (1926). Experimentelle untersuchungen über phänomenale Identität (Experimental investigations of phenomenal identity). *Psychologische Forschung*, 7, 81–13.
- 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.
- Troje, N.F. (2003). Reference frames for orientation anisotropies in face recognition and biological-motion perception. *Perception*, 32 (2), 201-10.
- Troje N F, Sadr J, Geyer H, Nakayama K,(2006). Adaptation aftereffects in the perception of gender from biological motion. *Journal of Vision* 6, 850 – 857.
- Troje, N.F., Westhoff, C. (2006). The inversion effect in biological motion perception: evidence for a "life detector"? *Curr Biol.*16(8),821-4.
- Troscianko T, Holmes A, Stillman J, Mirmehdi M, Wright D, Wilson A. (2004) What happens next? The predictability of natural behaviour viewed through CCTV cameras. *Perception.* 33(1):87-101.
- Todd, J.T. (1983). Perception of gait. *J Exp Psychol Hum PerceptPerform.* 9(1), 31-42.
- Unuma, M., K. Anjyo, & R. Takeuchi (1995). Fourier principles for emotion-based human figure animation, *Proceedings of ACM SIGGRAPH '95*, ACM Press, 91-96.
- Vaina, L.M., Solomon, J., Chowdhury, S., Sinha, P., & Belliveau, J.W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proc Natl Acad Sci U S A.* 98(20), 11656-61.
- Vaina, L.M.V., Beardsley, S.A., and Rushton, S. (2004). *Optic Flow and Beyond*. Kluwer Academic Press, Dordrecht, NL.
- Vallortigara, G., & Regolin, L. (2006). Gravity bias in the interpretation of biological motion by inexperienced chicks. *Curr Biol.* 16(8), R279-80.
- Vangeneugden, J, Pollick, F, & Vogels, R. (2009). Functional differentiation of macaque visual temporal cortical neurons using a parametric action space. *Cereb Cortex.* 19(3), 593-611.

- Vangeneugden, J., Vancleef, K., Jaeggli, T., Van Gool, L., & Vogels, R. (2010). Discrimination of locomotion direction in impoverished displays of walkers by macaque monkeys. *Journal of Vision*, 10, 22.1-22.19.
- Vangeneugden, J., De Mazière, 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. *J Neurosci*. 31(2), 385-401.
- Van Overwalle F, & Baetens K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage* 48(3), 564-584.
- Vanrie J, Verfaillie K. (2004). Perception of biological motion: a stimulus set of human point-light actions. *Behav Res Methods Instrum Comput.*, 36(4):625-629.
- Vanrie, J., Dekeyser, M., & Verfaillie, K. (2004). Bistability and biasing effects in the perception of ambiguous point-light walkers. *Perception*, 33(5), 547-560.
- Viviani, P., Stucchi, N. (1989). The effect of movement velocity on form perception: geometric illusions in dynamic displays. *Percept Psychophys*, 46(3), 266-274.
- Walk, R.D., & Homan, C.P. (1984). Emotion and dance in dynamic light displays. *Bull. Psychon. Soc.* 22, 437-440.
- Wang, L., & Jiang, Y. (2012). Life motion signals lengthen perceived temporal duration. *Proc Natl Acad Sci U S A*. 109(11), E673-7.
- Webb, J.A. & Aggarwal, J.K. (1982). Structure from motion of rigid and jointed objects. *Artif. Intell.* 19, 107-130.
- Wiley, D.J., & Hahn, J.K. (1997). Interpolation synthesis of articulated figure motion. *IEEE Computer Graphics and Applications*, 17(6), 39-45.
- Wertheimer, M. (1923). Laws of organization in perceptual forms. First published as *Untersuchungen zur Lehre von der Gestalt II*, in *Psychologische Forschung*, 4, 301-350.
- Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philos Trans R Soc Lond B Biol Sci*. 358(1431), 593-602.

Figure captions

- 1. Perceptual organization of simple motion displays:** **A** Induced motion (Duncker, 1929): While in reality the external frame moves and the dot is stationary, the dot is perceived as the moving element. (The following examples are taken from Johansson (1950)): **B** Three dots that move along straight lines are perceptually grouped into two pairs of dots that move up and down, with a periodic ‘contraction’ of their virtual connection line horizontally. **C** Two dots that move vertically and two that move along a circle are grouped into a single line that moves vertically. In addition, the exterior points are perceived as moving horizontally. **D** Two dots, where one moves along a straight line and the second along piecewise curved paths, is perceived as a ‘rotating wheel’, where one dot is rotating about the other.
- 2. Point-light biological motion stimulus.** **A** Light bulbs or markers are fixed to the major joints of a moving human. **B** Presentation of moving dots alone results in a point-light stimulus that induces the vivid perception of a moving human. (Reproduced from Giese, M.A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nat Rev Neurosci* , 4, p.180, with permission from Nature Publishing Group.)
- 3. Informative cues in body motion stimuli.** The global configuration of a human body can be recovered either from: **A**, local form features (e.g. orientation and positions of limbs or limb parts), or **B**, from local motion features, which specify for each time point a complex instantaneous optic flow field. **C** Trajectories of individual dots, like the ones of the feet, can also provide sufficient information for the solution of specific biological motion tasks, e.g. detection of walking direction. **D** Equivalent of a ‘life detector’ in the form domain: The direction of the nose in a scrambled face image (middle panel) makes it easy to determine the heading direction of the face (upper panel). This detection is more difficult if the picture is rotated upside down (‘inversion effect’).
- 4. Top-down effects in the processing of body motion.** **A** Visual search task for point-light walkers: The target is the walker walking to the left side. (Reproduced with permission from Cavanagh, P., Labianca, A.T., Thornton, I.M. (2001). Attention-based visual routines: sprites. *Cognition*, 80, p. 56, with permission from Elsevier.) **B** Stimulus demonstrating strong interference between shape recognition and body motion perception. (Reproduced from Hunt, A.R., & Halper, F. (2008). Disorganizing biological motion. *J Vis.* 8(9), 12, p. 3, with permission of the Association for Research in Vision and Ophthalmology.) **C** Motion stimulus by Fujimoto & Yagi (2008), showing that body motion processing interacts with the organization of ambiguous coherent motion of a grating. The background is preferentially perceived as moving in the direction that would be compatible with a forward locomotion of walker / runner. Similar observations hold for point-light patterns. (Figure modified from Fujimoto, K., Yagi,

A. (2005). Motion illusion in video images of human movement. In: F. Kishino et al. (eds.): *ICEC 2005*, LNCS 3711, Springer-Verlag, Berlin / Heidelberg, p. 532, reproduced with permission from Springer Verlag.)

5. **Models of body motion recognition.** **A** Example for a model for movement recognition by internal simulation of the underlying motor behavior. The core of the MOSAIC model by Wolpert et al. (2003) is a mixture of expert controllers for different motor behaviors, such as walking or kicking. Forward models for each individual controller predict the sensory signals that would be caused by the corresponding motor commands. These predictions are compared with the actual sensory input. The classification of observed movements is obtained by choosing the controller model that produces the smallest prediction error. **B** Neural architecture for body motion recognition, following models by Giese & Poggio (2003) and Fleischer et al. (2013). The model assumes processing in two parallel pathways that are specialized for form and motion features. Model neurons at different levels mimic properties of cortical neurons. Recognition in the form pathway is accomplished by integrating the information from sequences of recognized body shapes (recognized by ‘snapshot neurons’). Recognition from local motion features is accomplished by the detection of sequences of characteristic optic flow patterns. Recognition is first accomplished in a view-specific manner within view-specific modules. Only at the highest hierarchy the outputs of these view-specific modules are combined, achieving view-independent recognition. (Potentially relevant cortical areas in monkey and human cortex are indicated by the abbreviations below the modules of the model. See above references for further details.)

Figure 1

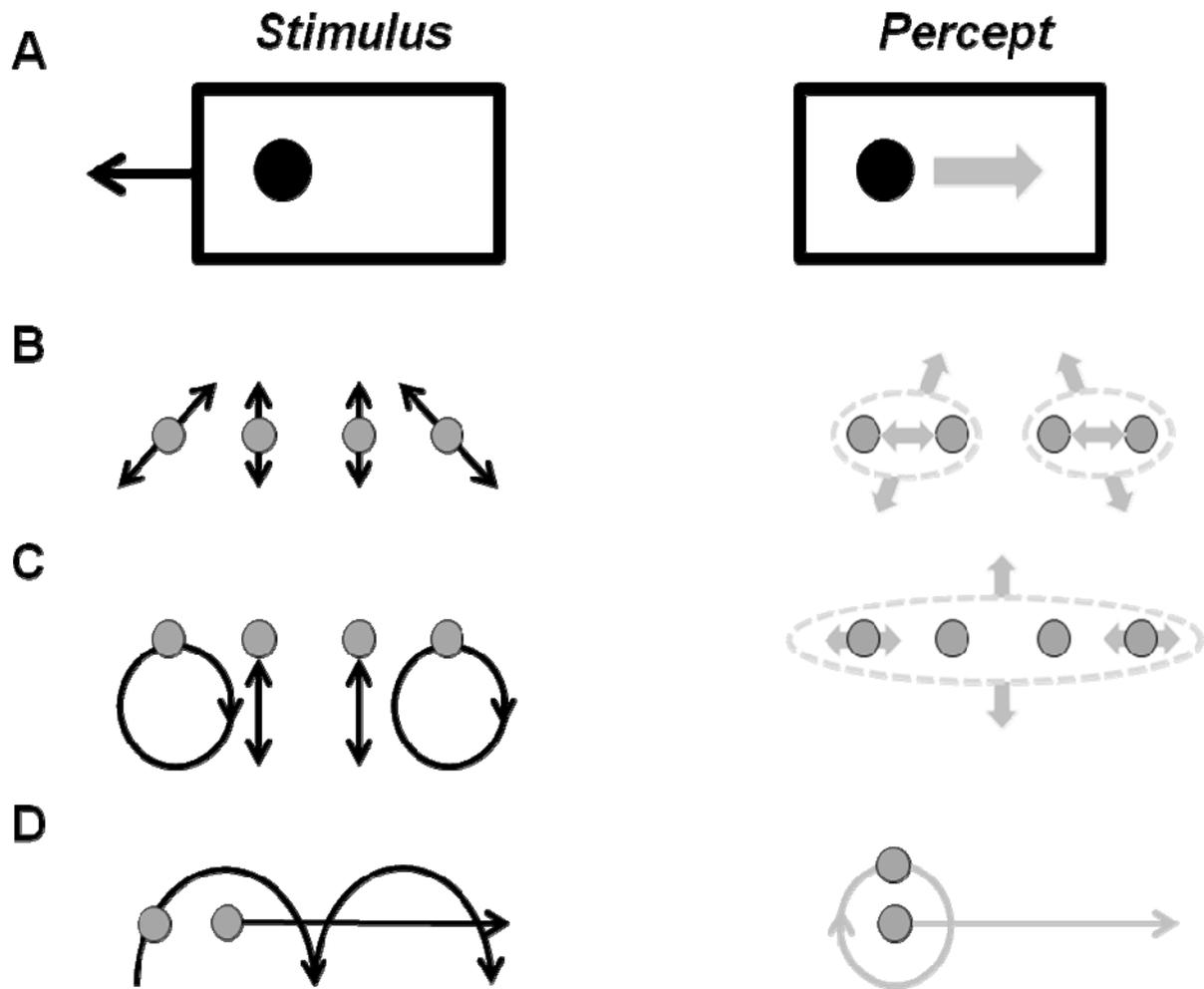
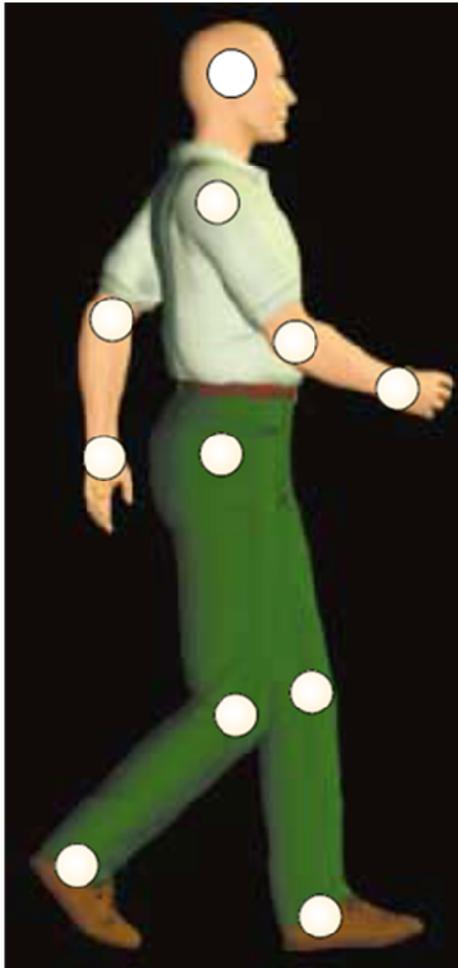


Figure 2

A



B

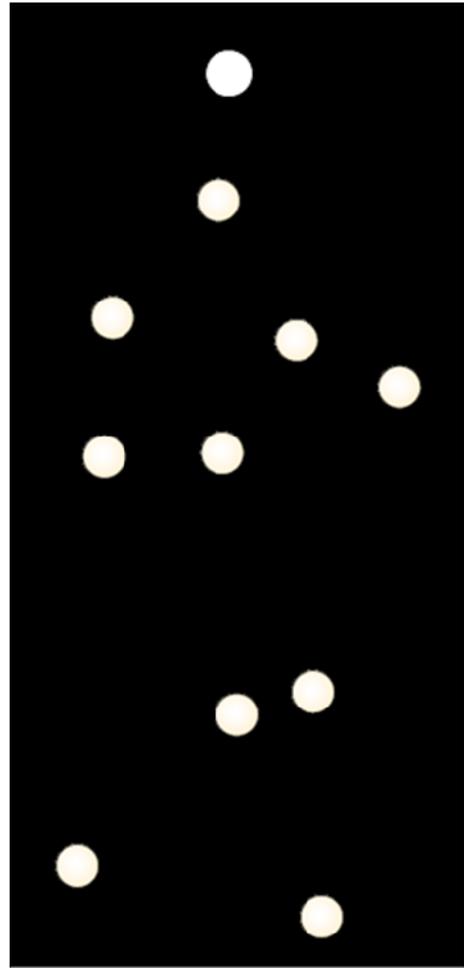


Figure 3

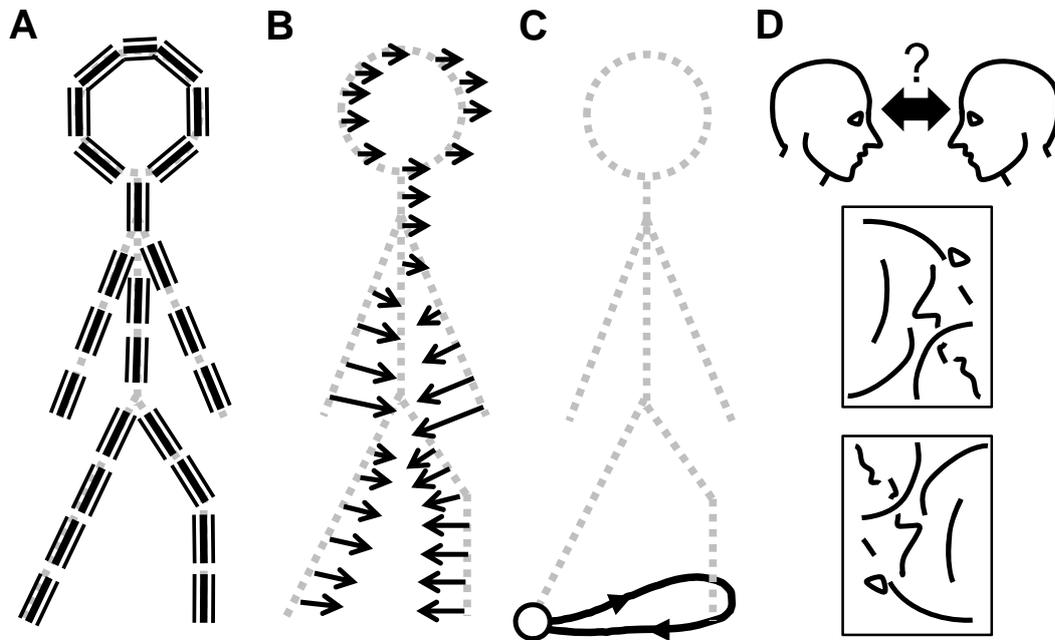


Figure 4

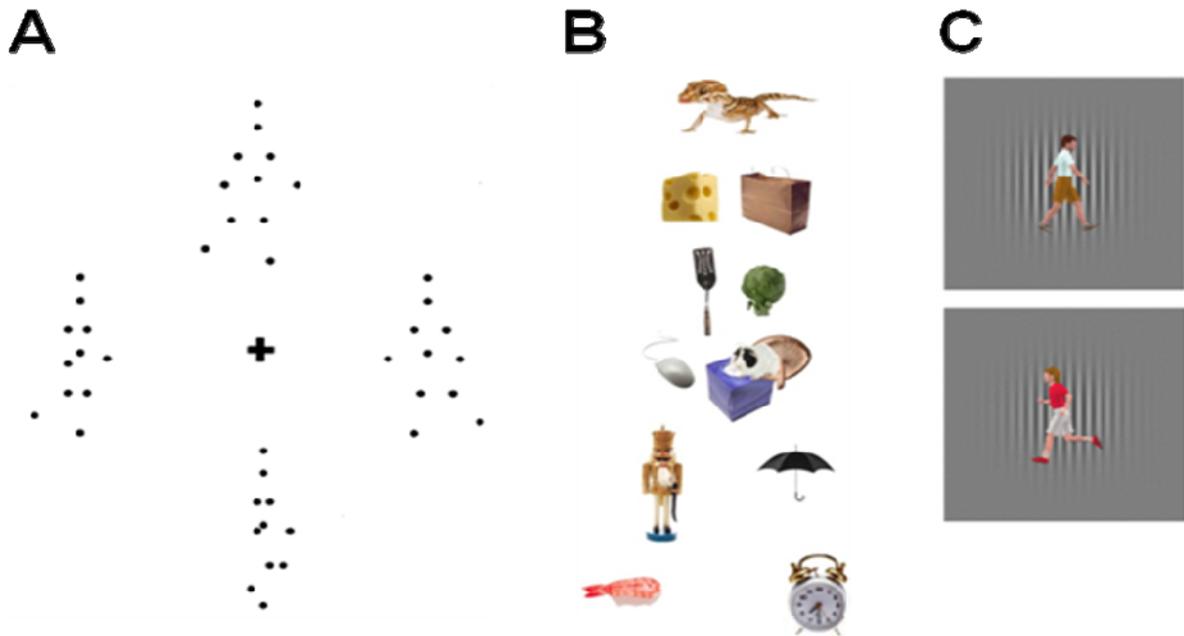


Figure 5

