UCLA
UCLA Electronic Theses and Dissertations

Title
Biological motion perception: action interaction and prediction

Permalink
https://escholarship.org/uc/item/2nz129mf

Author
Su, Junzhu

Publication Date
2016

UNIVERSITY OF CALIFORNIA

Los Angeles

Biological Motion Perception:

Action Interaction and Prediction

A dissertation submitted in partial satisfaction of the requirements for the

requirements for the degree Doctor of Philosophy

in Psychology

by

Junzhu Su

2016

ABSTRACT OF THE DISSERTATION

Biological Motion Perception:

Action Interaction and Prediction

by

Junzhu Su

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2016

Professor Hongjing Lu, Chair

Humans, as social creatures, are especially adept in perceiving others' actions and

inferring mental states such as intention based on non-verbal cues such as body movements. In

recent years, substantial advances have been achieved in our understanding of biological motion

perception and its relation to other cognitive and motorsensory processes, as well as the

underlying neural correlates and computational components. In my thesis, I first focused on the

top-down influence of meaningful interactive actions on stimulus visibility. We found that

meaningful interaction boosted the visibility of single actor in binocular rivalry paradigm.

Second, I investigated the action prediction modulated by action interaction (i.e. either single

actor or two interactive actors) with temporal occluder paradigm. We found that observers

obtained higher temporal precision in single actor condition but better sensitivity to posture

sequence or rotation in two interactive actor condition. Third, we designed novel paradigm by

briefly flashing joints near the moving limb to determine how action representation interacts with generic mechanisms for localization of moving features to bias perceived positions of moving joints. We found that global posture information (inversion or upside-down) modulated the effect size of well-established visual illusion (i.e. flash lag effect). Last but not least, I used this paradigm to dissociate local and global mechanisms that may underlie action prediction. By briefly flashing the stationary skeleton during the action viewing, either aligned or misaligned with the walking posture, we measured the motion-induced posture change. We demonstrated the existence of a global mechanism for anticipating changes in whole-body posture over time. However, although inversion of body orientation weakens global processing in biological motion perception, it does not eliminate the use of the global mechanism in predicting future posture even when the action is presented upside-down. This finding is consistent with previous studies of the inversion effect in biological motion perception (Pavlova & Sokolov, 2000; Sumi, 1984), and highlights the robust use of the global mechanism in predicting future actions.

The dissertation of Junzhu Su is approved.

Martin M. Monti

Mirella Deparetto

Dario L. Ringach

Hongjing Lu, Committee Chair

University of California, Los Angeles

2016

To my family,

Luling Qin and Dong su,

Whose love made me as today.

To my lab,

Hongjing Lu, and the current and previous colleagues,

Whose knowledge, understanding and help

Led me through the way.

To my other half,

James An,

Whose appearance lighted up my life,

And completed my study with a happy ending.

Table of Contents

List of Figures

VITA

EDUCATION

Peking University Department of Psychology 2011

    M.S., Cognitive Neuroscience

    Thesis: "Neural correlates of perceptual learning of face view, identity and gender"

Peking University Department of Psychology 2008

    B.S., Basic Psychology

    Thesis: "Unconscious Ebbinghaus Illusion"

CONFERENCE/POSTER PRESENTATIONS

Su, J., & Lu, H. (2015, May). *Action Prediction and the Flash-Lag Effect*. Poster Presented at 15th Vision Sciences Society, St. Pete Beach, FL.

van Boxtel, J., Su, J., & Lu, H. (2014, December). *Social interactions receive priority access to conscious perception*. Talk Presented at 10th Australasian Society for Cognitive Science Conference, Melbourne, Australia

Su, J., van Boxtel, J., & Lu, H. (2014, May). *Meaningful actions and interactions receive priority in conscious perception*. Poster Presented at 14th Vision Sciences Society, St. Pete Beach, FL.

Su, J., van Boxtel, J., & Lu, H. (2013, May). *Action prediction and interaction enhances stimulus visibility during binocular rivalry*. Poster Presented at the 13th Vision Sciences Society, Naples, FL.

Su, J., van Boxtel, J., & Lu, H. (2012, May). *'What' and 'When' in Action Prediction*. Poster Presented at the 12nd Vision Sciences Society, Naples, FL.

Fang, F., Su, J., Chen, C., & He, D. (2011, September). *Perceptual learning induces fast processing and efficient representation of face*. Poster Presented at the 11st Vision Sciences Society , Naples, FL.

Su, J., Chen, C., He, D., & Fang, F. (October, 2010). *Reduced N170 latency and Gamma power after perceptual learning of face view discrimination*. Poster Presented at the 7th International Conference on Cognitive Science, Beijing, China.

Su, J., & Fang, F. (September, 2009). *Unconscious Ebbinghaus Illusion*. Poster presented at

Workshop on Cognitive Science from Cellular Mechanisms to Computational Theories, Beijing, China.

PUBLICATIONS

Van Boxtel, J. J., Peng, Y., Su, J., & Lu, H. (in revision). Individual Differences in High-level Biological Motion Tasks Correlate with Autistic Traits. *Vision Research*

Su, J., & Lu, H. (in revision). Flash-lag effects in biological motion. *Vision Research*

Su, J., van Boxtel, J. J., & Lu, H. (2016). Social Interactions Receive Priority to Conscious Perception. *PloS one, 11(8),* e0160468. PMID: 27509028

Su, J.[1], Tan, Q.[1], and Fang, F. (2013). Neural correlates of face gender discrimination learning. *Experimental Brain Research, 225(4),* 569- 578. PMID: 23307160

Su, J., Chen, C., He, D., & Fang, F. (2012) Effects of face view discrimination learning on N170 latency and amplitude. *Vision Research, 61,*125-131. PMID: 21911001

Bi, T., Su, J., Chen, J., & Fang, F. (2009) The role of gaze direction in face viewpoint aftereffect. *Vision Research, 49,* 2322 - 2327. PMID: 19607854

Chapter 1 Introduction

Biological motion is a term used by vision scientists to refer to the perception of a moving, animate object. Often, the stimuli used in biological motion experiments are just a few moving dots that reflect the motion of some key joints of the moving organism. Before Johansson's (1973) work, researchers on visual motion perception has dealt mostly with simple patterns of mechanical motions. As a rule, rigid objects in rotary or translational motion have been chosen as the distal stimuli, and the proximal stimulus patterns investigated have been projections of such objects in motion. Stimulus patterns representing living animals in motion had rarely been studied (Michotte's (1963) study of perception of larva motion is as an important exception.) The relative lack of studies concerning visual perception of animal and human motion patterns does not adequately reflect the biological importance of this type of visual perception. It seems evident that, throughout animal evolution, valid information about other animals' motion would have had a very high survival value.

Our everyday experience also tells us that human vision not only detects motion directions of humans and animals, but also distinguishes different types of motion patterns. We immediately see whether a person is walking, running, or dancing, and also if he is moving forward with identical speed in these three cases. It is also a common experience that our visual apparatus is very sensitive to small deviations from such standard patterns. We immediately recognize, for instance, a slight limp in walking; we distinguish between a tired and an elastic gait, etc. Furthermore, humans can recognize a person exclusively from his style of walking, and his gestures, etc. The geometric structures of body motion patterns in humans and other animals (e.g., the patterns of walking) are determined by the construction of their skeletons. Human walking, for instance, as well as the same types of motion in most domestic animals, can readily be described as combinations of several pendulum-like motions of the extremities relative to a joint, combined with an elastic bending of the spinal column. Different types of human body movements, such as walking, running, dancing, and skating, are all built up from such combinations of pendulum-like motions, which are highly specific for the different types of motions. The typical pendulum motions about the hip and the knee joints during walking of a young man and their phase relations are shown in Figure 1-1.
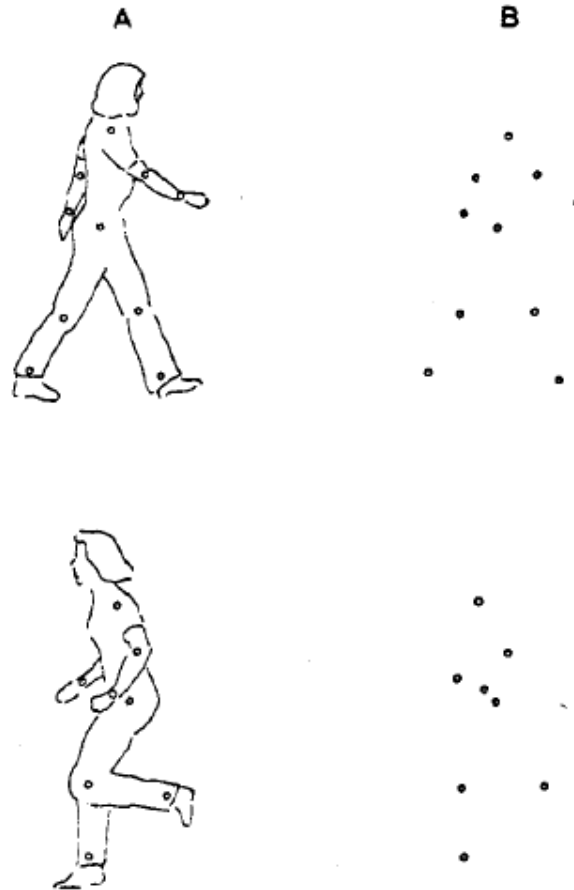
Figure 1-1 (A) Outline contours of a walking and a running subject; (B) the corresponding dot configuration (Johansson, 1973).

In order to better isolate biological motion from other bodily features, a tool has been created by reducing rendered video actions to rarefied stimuli called Point-Light Displays (PLDs), in which the moving person is completely invisible except for about a dozen strategically placed points of light at the major joints. PLDs are not identified when lights are stationary but are quickly recognized as human actions when they are moving. This stimulus was first invented by Johansson (1973) as a minimal information display to demonstrate that indeed the visual system can utilize motion information, with no further cues, to infer the correct structure and motion of a human being. More specifically, small light bulbs are attached to a subject's body at each of its joints (e.g., ankle, knee, hip, shoulder, elbow, wrist). The subject is then placed in a dark room and filmed while performing various activities. Single frames of the resulting film look to naive observers merely to be pictures of a few randomly placed dots. But when the film is shown at normal speeds, naive observers almost immediately (within 100-

1000ms) see the dots as a person walking, running, etc. In fact, the perception is so powerful that it is impossible to force oneself to interpret the dots in any other manner.

Although biological motion generally includes the kinematics particular to both animal and human bodies, in this thesis, I mainly focused on the latter category. Human biological motion is also simply termed "action", containing rich social information that is crucial for the ability to perceive what others are doing, and to infer from gestures and expressions what others may be intending to do. Researchers have found that the human visual system is highly sensitive to biomotion. From only a few moving point lights, attached to the joints of an otherwise invisible moving actor, people readily perceive the underlying human figure, categorize the displayed action after viewing it for only fractions of a second (Jansson & Johansso.G, 1973), and can even perform subtle tasks such as gender recognition (Kozlowski & Cutting, 1977; Sumi, 2000; Troje, 2002) or automatically extract information of actors' identity (Troje, Westhoff, & Lavrov, 2005), emotion (Clarke, Bradshaw, Field, Hampson, & Rose, 2005), and intention (Saxe, Xiao, Kovacs, Perrett, & Kanwisher, 2004; Zacks et al., 2001) within only milliseconds with little conscious effort (Blake & Shiffrar, 2007). This broad range of abilities, jointly referred to as biological motion perception, is remarkable given the sparse distribution of points and the many degrees of freedom involved.

Recent literature has investigated the question of how relevant information is encoded in biomotion patterns and how such information can be retrieved (Blake & Shiffrar, 2007). We ask: What are the underlying cognitive processes involved in action inference (i.e., action prediction or action interaction)? What role do top-down mechanisms play in action inference?

1.1 Action recognition/representation

1.1.1 Local mechanisms: optical flow information

Experiments have found that (much like face perception, which also contains rich social information) biological motion perception exhibits substantial inversion effects: observers can recognize upright actors more easily (higher accuracy and shorter reaction time) than inverted actors (Dittrich, 1993; Proffitt & Bertenthal, 1990; Reed, Stone, Bozova, & Tanaka, 2003). Early theories claimed that the strongly impaired perception of inverted (upside-down) biological-motion point-light displays is due to impaired configural processing in a highly trained expert system. However, Troje and Westhoff (2006) found that observers can readily retrieve local

information about direction from scrambled point-light displays of humans and animals. In their study, participants were asked to indicate in which direction the animals were walking while they viewed point-light displays that were inverted and/or scrambled in different ways. Surprisingly, they found that adult humans can readily judge direction of motion from most scrambled point-light displays. This result shows that the overall configuration of dots on the body is not important. Further, when only parts of the display were inverted, the participants' judgments were correct as long as the dots associated with the local motion of the feet remained intact and located at the bottom of the dot array, indicating a visual filtering mechanism tuned to the characteristic motion of animal limbs in locomotion, which serves as a general detection system for the presence of articulated terrestrial animals: a 'life-detector'.

Moreover, Hirai and his colleagues (2011) also found that local cues, by themselves, can almost unambiguously indicate the facing direction of the agent in biomotion stimuli, although the local motion information tends to interact with the global configural information to some extent. They found that when the walkers are scrambled (i.e., containing no configural cues to the direction in which a walker was facing), manipulating the relative vertical location of the walker's feet significantly affected observers' performance on a direction-discrimination task. For the interaction between local and global information, they found that among three levels of global organization (the vertical location of the individual dots: upright, inverted, or scrambled), in both the globally upright and globally inverted conditions, observers' accuracy decreased when the local trajectories of displays were inverted. The local inversion effect was stronger in the globally upright condition when the walker was presented with its feet located in their familiar position below the rest of the body, whereas in the globally inverted condition when the global organization of the stimulus was inverted so that the walker's feet were located at the top of the display, differences between the two local-motion conditions were virtually eliminated. These results indicated that direction discrimination is largely due to local motion signals, but also is influenced to a certain extent by the global information. Besides the facing direction discrimination, Chang and colleagues (2008) found that animacy ratings of scrambled displays that contain solely local trajectory cues can be reliably elicited within exposure times as short as 200 ms, and are not only orientation-specific but also linearly correlated with the ability to discriminate direction of walking. This finding indicates that the mechanisms responsible for

processing local biological motion signals not only retrieve locomotive direction but also aid in assessing the presence of animate agents in the visual environment.

1.1.2 Global mechanisms: body configural information

Based on some behavioral research as mentioned above, biomotion perception may rest on image motion signals such that vector analysis of component of body parts (defined by pairs of dots) is incorporated into a single structured percept (structure form motion). Other studies have demonstrated that patients with lesions to motion processing areas of the dorsal stream or with profound spatial deficits (e.g., Williams syndrome patients) who are severely impaired in image motion perception can still easily perceive biological motion (Jordan, Reiss, Hoffman, & Landau, 2002; McLeod, Dittrich, Driver, Perrett, & Zihl, 1996; Vaina, Lemay, Bienfang, Choi, & Nakayama, 1990). These findings suggest that biological motion perception does not solely rely on the analysis of image motion signals.

In addition to evidence from patient studies, studies using limited lifetime technique showed that subjects can spontaneously recognize a moving human figure in displays without fixed local image motion trajectory, with performance very similar to that for classic PLDs. More specifically, in Beintema and Lappe's study (2002), instead of using the classic point-light displays in which each image frame has light points attached to the joints of a moving human figure and the motion signal is carried by the apparent image motion of each individual point in two successive animation frames, they used the stimuli in limited lifetime, which dissociated the information of point positions of the body and the motion of these points over time. In the limited lifetime stimulus, eight light points were positioned not on the joints, but rather on the limbs at a random position between joints with equal probability (Figure 1-2). Therefore, no individual point carried the valid image motion signal of the limb movement. But as each animation frame displays a random-dot sampling of a static posture of the body, the sequence of these static postures also carries information about the form and motion of the body. They called this stimulus the "sequential position walker". Using these newly created stimuli, they found that subjects can spontaneously recognize or discriminate the walking direction or the coherence of a walking figure in the absence of image motion. Thus, although image motion may generally aid processes such as segmenting figure from background, they suggested that it is not the basis for the percept of biological motion. Rather, biological motion may be derived from dynamic form

information on body posture evolving over time. Beintema and colleagues (2006) manipulated the amount of local image motion based on this limited lifetime technique and found that irrespective of the frame rate, performance of detection or discrimination did not vary with point lifetime, indicating that the analysis of biological motion does not benefit from examining local motion. The reliance of perception on the number of displayed points and frames within a certain duration of time further supports the idea that biological motion is perceived from a sequence of spatiotemporally sampled forms.
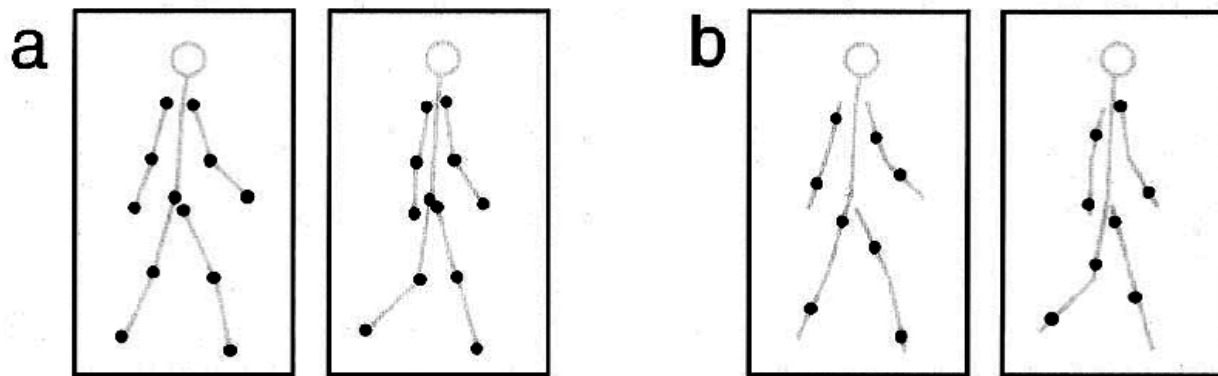
Figure 1-2. Standard biological motion stimuli (a) consist of a frame animation of the motion of light points attached to the joints of a moving human figure. In the sequential position stimulus (b), light points were positioned anywhere on the limbs and jumped to another randomly selected position for each frame. (Beintema & Lappe, 2002)

1.2 Biologically-inspired model for action recognition

1.2.1 Feedforward hierarchical model combing both form and motion pathways

To account for numerous empirical findings of biological motion perception, Giese and Poggio (2003) suggested using a neurophysiologically plausible and quantitative model as a tool to organize and interpret the experimental data. This model includes two parallel processing streams: the ventral stream that is specialized for the analysis of form, and the dorsal stream that is specialized for optic-flow/motion information. Both pathways comprise hierarchies of neural feature detectors that extract form or optic-flow features with increasing complexity and position or size invariance along the hierarchy (Figure 1-3).

For the form pathway, the first level consists of simple cells modeled as local orientation detectors or gabor-like filters, with eight preferred orientations, two spatial scales that differ by a factor of two, and the receptive field sizes within the range of neurons in monkey V1 (Dow, Snyder, Vautin, & Bauer, 1981). The next level of the form pathway contains position- and scale-invariant bar detectors corresponds to complex-like cells in area V1 (Hubel & Wiesel, 1962), or to neurons that are increasingly invariant to position changes in areas V2 and V4 (Gallant, Braun, & Vanessen, 1993; Hegde & Van Essen, 2000), which extract local orientation information from first level. A neurophysiologically plausible mechanism for achieving position and scale invariance is the pooling of the responses of neurons with similar preferred orientations, but with different receptive field positions and spatial scales (Fukushima, 1980; Perrett & Oram, 1993). The model assumes that this pooling is accomplished by a nonlinear maximum-like operation rather than by linear summation (Mel & Fiser, 2000; Riesenhuber & Poggio, 1999), as demonstrated by the complex cells in the visual cortex of cats (Lampl, Ferster, Poggio, & Riesenhuber, 2004) and neurons in area V4 of macaques (Gawne & Martin, 2002). The next level of the form pathway contains snapshot neurons that are selective, for instance, for body shapes. These model neurons are similar to view-tuned neurons in monkey inferotemporal cortex (area IT) (Logothetis & Sheinberg, 1996), and in the STS of monkeys and humans (Grossman & Blake, 2002; Perrett et al., 1985). These neurons have large receptive fields (>8°), show substantial position- and scale-invariance, and can become tuned to complex shapes through learning (Logothetis, Pauls, & Poggio, 1995). Activity that is selective specifically for human body shapes has also been found in the human lateral occipital complex (Downing, Jiang, Shuman, & Kanwisher, 2001), occipital and fusiform face areas (Grossman & Blake, 2002). Modeled by Gaussian radial basis functions, these neurons receive inputs from the invariant bar detectors on the previous hierarchy level and then adjust the centers of the basis functions during training so that each snapshot neuron encodes one key frame from a training sequence. However, the model does not address how an optimum set of key frames can be learned automatically. The highest hierarchy level of the form pathway consists of motion pattern neurons. These model neurons temporally smooth and summate the activity of all snapshot neurons that contribute to the encoding of the same movement pattern. According to physiological data, motion pattern neurons in monkey and human cortex are probably located in the STS (Grossman & Blake, 2002; Perrett et al., 1985) and the premotor cortex (area F5) (Buccino et al., 2001; Rizzolatti, Fogassi,

& Gallese, 2001), and possible also the fusiform and occipital face areas (Grossman & Blake, 2002).
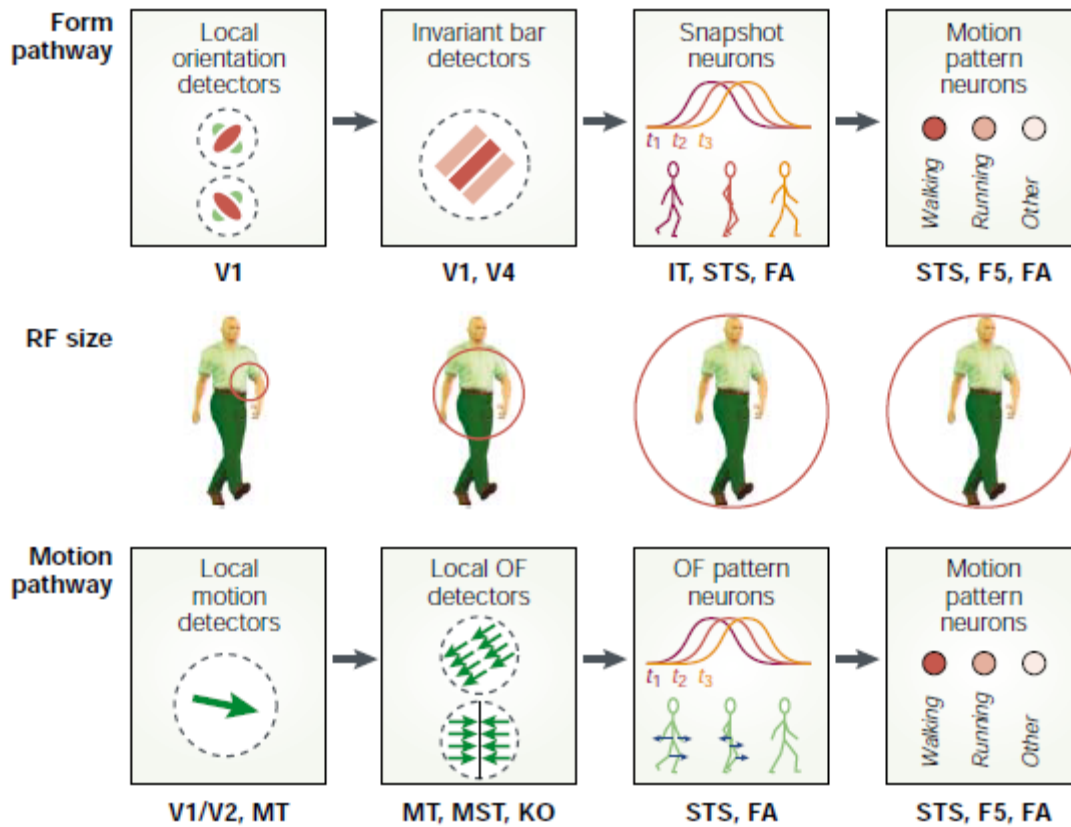
Figure 1-3. Hierarchical neural model of human action perception with two pathways for the processing of form and motion (optic flow). The approximate size of the receptive fields compared to typical stimuli is indicated in the middle row. Insets show the different types of neural detectors at different levels of the hierarchy. IT, inferotemporal cortex; KO, kinetic occipital cortex; OF, optic flow; RF, receptive field; STS, superior temporal sulcus; V1, primary visual cortex. (Giese & Poggio, 2003)

For the motion pathway, the first level of the motion pathway consists of local motion detectors that correspond to direction-selective neurons in V1 (Smith & Snowden, 1994) and component motion-selective neurons in area MT (Rodman & Albright, 1989). Their equivalent receptive field sizes are in the range of direction-selective neurons in V1, and of foveal neurons in area MT (Albright & Desimone, 1987; Dow et al., 1981). The second level of the motion pathway consists of neurons with larger receptive fields that analyze the local structure of the optic-flow fields induced by movement stimuli. There are two types of local optic flow detector. The first is selective for translation flow and corresponds to motion pattern neurons in area MT

(Rodman & Albright, 1989) with low or bandpass tuning with respect to speed (Lagae, Raiguel, & Orban, 1993), and has direction tuning curves with a width of about 90° (Smith & Snowden, 1994). The second class of local optic-flow detectors is selective for motion edges (horizontal and vertical). Their output signals are computed by combining the responses of two adjacent subfields with opposite direction preferences (Figure 1-3) in a multiplicative way. Neurons with such opponent motion selectivity have been found in several areas in the dorsal processing stream, including areas MT (Xiao, Raiguel, Marcar, Koenderink, & Orban, 1995), MST dorsal (Tanaka & Saito, 1989) and MST lateral (Eifuku & Wurtz, 1998). They are probably also present in the kinetic occipital area (KO) in humans (Orban et al., 1995). The optic-flow pattern neurons on the third level of the motion pathway are equivalent to the snapshot neurons in the form pathway, which are selective for complex optic flow patterns that arise for individual moments of biological movement patterns. Like the snapshot neurons, the motion pattern neurons are modeled by Gaussian radial basis functions that receive their inputs from the previous hierarchy level. Such optic-flow pattern neurons might be found at different locations in the visual cortex, in particular in the STS (Grossman & Blake, 2002; Perrett et al., 1985), fusiform and occipital face areas (Grossman & Blake, 2002), and maybe area MST.

Generally, this hierarchical model can account for results of many studies, but it has several flaws. (1) *Fixed parameters of "snapshots"*. A central postulate of this model is that a set of learned patterns are encoded as sequences of 'snapshots' of body shapes in the form pathway, and by sequences of complex optic flow patterns in the motion pathway. This assumption may not hold, since newborn humans show preference for human action even without any learning experience (Bardi, Regolin, & Simion, 2011; Bertenthal, Proffitt, & Cutting, 1984; Fox & Mcdaniel, 1982; Simion, Regolin, & Bulf, 2008). (2) *Single modality*. The model includes only areas that are primarily visual (Figure 1-3) in movement recognition. Evidence actually showed that multimodal signals can also activate STSp, as when one hears footstep sounds produced by people walking compared to noise (Bidet-Caulet, Voisin, Bertrand, & Fonlupt, 2005), or when a monkey hears species-specific, emotionally charged vocalizations (Gil-Da-Costa et al., 2004), indicating that portions of STSp receive multimodal input related to human activity recognition (Wheaton, Thompson, Syngeniotis, Abbott, & Puce, 2004). (3) *Over-simplification of the two parallel pathways*. In the model, the pathways specialized for the analysis of form and motion (optic flow) information, are totally separate. In monkey and human

brains, the two processing streams interact at several levels. Such couplings, for example at the level of the STS, can be easily integrated into the model and improve its recognition performance without changing the basic principles. (4) *Lack of top-down influence*. The model assumes that in its basic, initial operation (akin to 'immediate recognition'), the hierarchy in both pathways is predominantly feedforward (apart from local feedback loops). Although good recognition performance can be achieved in most cases for fast recognition without the need of top- down signals, feedback signals can help for longer stimulus presentations, or be initiated by attention modulation or task instructions. Thus, this model should have included the mirror neural system and top-down modulation related to biological motion perception to be more comprehensive.

1.2.2 Mirror Neuron System: common coding for observation and execution

According to the common coding theory (Prinz, 1992; Prinz, 1997), perceived events and planned actions share a common representational domain, that is, human brains utilize the same coding mechanisms when viewing actions conducted by other people and conducting the same action by themselves. The two assumptions of the common coding theory are: (1) that event codes and action codes are considered as the functional basis of percepts and action plans, respectively; (2) that they share the same representational domain and are therefore commensurate (Figure 1-4).
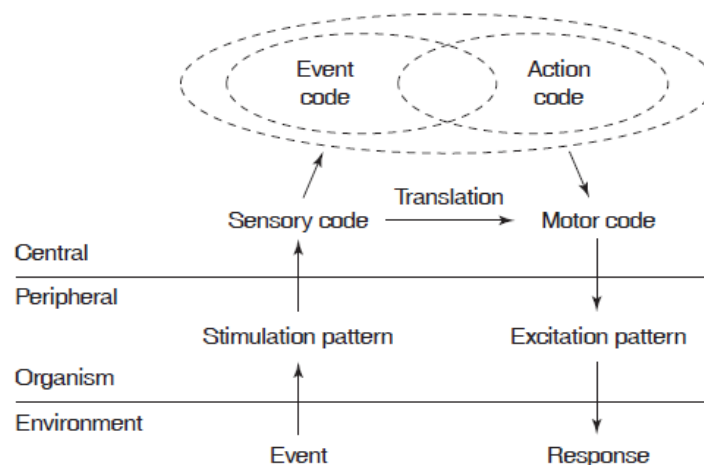
Figure 1-4. Major functional components that underlie perception and action control. On the left-hand side (upward arrows), events in the environment lead to patterns of stimulation in the sense organs (peripheral) and generate sensory codes in the brain

(central). On the right-hand side, the activity travels down, from motor codes to patterns of excitation in the muscles to the action (response). (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996)

This was first revealed in monkey studies, when the roles played by mirror neuron system were investigated through animal-object interaction. Gallese and colleagues (1996) recorded electrical activity from 532 neurons in the rostral part of inferior area 6 (area F5) of two macaque monkeys. They discovered that a subset of F5 neurons ('mirror neurons', n = 92) not only discharged during goal-directed hand and mouth movements, but were also active when the monkey observed a similar action performed by the experimenter. However, in order to be visually triggered, those mirror neurons required an interaction between the agent of the action and the object of it. The sight of the agent alone or of the object alone (three-dimensional objects, food) was ineffective. Hands and mouth were by far the most effective agents. The actions most represented among those activating mirror neurons were grasping, manipulating and placing. In some of the mirror neurons the congruence had to be very strict (e.g., precision grip), indicating the possible role of this system in action recognition and, given the proposed homology between F5 and human Brocca's region, raises the possibility that it is involved in recognition of actions as well as phonetic gestures.

Further evidence supporting this common coding theory can be found in other studies. Evidence shows that motor laws that constrain one's performance also constrain one's perception of others' actions. These motor laws can be the two-thirds law (Dayan et al., 2007; Flach, Knoblich, & Prinz, 2004) or Fitts' law (Decety & Jeannerod, 1995; Grosjean, Shiffrar, & Knoblich, 2007). Similarly, neurological impairments affect a person's performance and action perception in the same way. Eskenazi and colleagues (2009) found that patient DS, who suffers from a frontal brain lesion, ignored target size not only when performing movements by himself but also when asked to judge whether others can perform the same movements, supporting the assumption of motor contributions (either action production or motor imagery) to action perception or the so called "common coding".

Jokisch and colleagues (2006) investigated the influence of viewing angles (frontal view, half-profile view, or profile view) on identity recognition of one's own person and familiar individuals such as friends or colleagues from walking patterns. Observers were requested to assign the person's name to the individual gait pattern. They found that whereas recognition

performance of one's own walking patterns was viewpoint independent, recognition rate for other familiar individuals was better for frontal and half-profile view than for profile view, supporting common coding theory since we possess better motor representation of our own action than of others' action. Sensorimotor system can also influence the emotional processing of biological motion. Sevdalis and Keller (2012) presented participants point-light displays of 200~1000ms duration, as well as static displays, of expressive and inexpressive dance performances. Subjects were asked to identify the intended expression intensity of the performer. Results showed that expression intensity could be discerned reliably only from dynamic displays, even as short as 200ms, though the accuracy of judgments increased with exposure duration. More interestingly, the performance accuracy was correlated with indices of informal music and dance experience, indicating the linkage between sensorimotor and cognitive-emotional processes underlying action understanding and social cognition. A brain imagining study (Grezes, Frith, & Passingham, 2004) showed that when subjects watched videos of themselves and of others lifting a box, and judged the beliefs of the actors about the weight of the box, a parietal premotor circuit was recruited during action perception, and the activity started earlier when making judgments about one's own actions as opposed to those of others. This finding indicates that when one observes one's own actions, there is a closer match between the simulated and perceived action than there is when one observes the actions of others. When the observers judged the actions to reflect a false belief, there was activation in the superior temporal sulcus, orbitofrontal, paracingulate cortex and cerebellum, reflecting a mismatch between the perceived action and the predicted action's outcomes derived from simulation.

However, some studies suggest that there are not yet conclusive arguments for a clear neurophysiological substrate supporting a common coding between perception and action, other than the premotor cortex (Decety & Grezes, 1999). In the review, Decety and Grezes categorized neuroimaging studies about action observation into three types: (1) "vision for action", when subjects were instructed to observe the actions so that they could imitate them immediately after the scanning session; (2) "vision for perception", when subjects were not told that they have to imitate these actions later on and were supposedly imitating the action implicitly; and (3) "vision for recognition", when subjects were explicitly required to view actions for the purpose of recognition (i.e., subjects were aware that they would be given a recognition memory test following the observation phase, which requires memory encoding). They concluded that the

strong conclusion that the neural substrate for action planning is activated during perception of action holds true only when the goal is to imitate that action (i.e., vision for action), when dorsal pathway is mostly activated (Figure 1-5A). But the neural substrates underlying the action-perception linkage are less clearly defined when the observer has a goal other than imitation in mind (i.e. s vision for recognition), for which the ventral pathway is mostly activated (Figure 1-5C). This separation indicated that when perception has an explicit goal, the data are consistent with the functional segregation of the labor in the visual pathways (see Figure 1-5A and C). However, when perception has no explicit aim, such as in the vision for perception tasks (Figure 1-5B), both visual pathways seem to be involved. Thus, the roles of the two pathways are more easily understood when considered from the point of view of the output (top-down processing) as suggested by Milner and Goodale (1995).
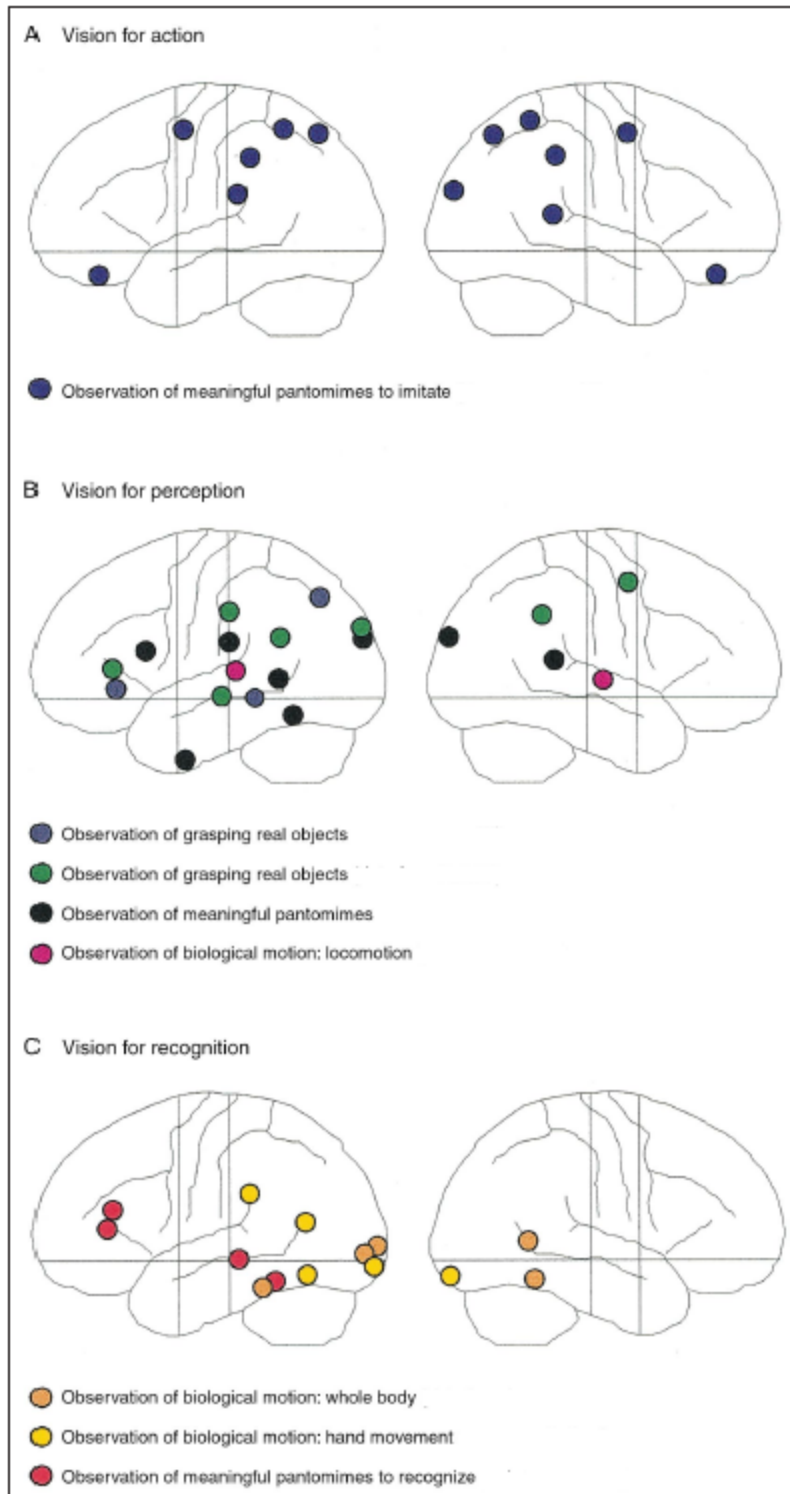
Figure 1-5. A summary of the results of neuroimaging studies during perception of action: vision for action (A); vision for perception (B); and vision for recognition (C). Activation foci are shown on a schematic brain registered to Talairach coordinates. For the sake of clarity activations found consistently in the V5 complex are not shown. (Gallese et al., 1996)

1.3 Action inference

1.3.1 Action interaction

The ability to coordinate our actions with those of others is crucial for our success as individuals and as a species. Progress in understanding the cognitive and neural processes involved in joint action has been slow and sparse, because vision scientists and cognitive neuroscientists have predominantly studied individual actions in isolation. However, in recent years, some advancements have been made by investigating perception and action in social context.

Interactive information can boost detectability of acting agent. The ability to interpret and predict other people's actions is highly evolved in humans and is believed to play a central role in their cognitive behavior. To investigate whether the ability involved in action interaction confers a tangible benefit to sensory processing, Neri and colleagues (2006) conducted an experiment to quantitatively measure whether visual discrimination of a human agent is influenced by the presence of a second agent. They found that the effect depended on whether the two agents interacted (by fighting or dancing) in a meaningful synchronized fashion that allowed the actions of one agent to serve as predictors for the expected actions of the other agent, even though synchronization was irrelevant to the visual discrimination task. These results demonstrated that action understanding has a pervasive impact on the human ability to extract visual information from the actions of other humans, providing quantitative evidence of its significance for sensory performance. Manera and colleagues (2011) investigated whether the effects of interpersonal action inference extends to activities in which no physical contingency is implied between the movements of the interacting individuals, such as communicative interactions. In their study, they asked subjects to observe point-light displays of two agents (A and B) performing separate actions. In the communicative condition, the action performed by agent B responded to a communicative gesture performed by agent A. In the individual condition, agent A's communicative action was substituted with a non-communicative action. Using a simultaneous masking detection task, they demonstrated that observing the communicative gesture performed by agent A enhanced visual discrimination of agent B, indicating that the communicative gestures of one agent can serve as a predictor for the expected

actions of the respondent, even if no physical contact between agents is implied. In summary, these studies show that interactive actor can enhance perceptual visibility of biological motion.

Social interactive information can enhance sensorimotor simulation of an anticipated or observed action. In a recent EEG study, to assess the effects of social interaction on action simulation, Kourtis and colleagues (2010) recruited three participants in their study: two of them performed both individual actions (e.g., lifting an object) and joint actions (e.g., passing the object), while the third participant ("loner") performed only individual actions. Their results showed that a person's motor system was more active when anticipating observing an individual action performed by the interaction partner than when anticipating observing the same action performed by a single actor, suggesting that interpersonal coordination between an actor and an observer affects the extent to which predictive action simulation takes place. Moreover, Hogeveen and Obhi (2012) reported greater activation of motor resonance (MR) during observation of recorded human actions when the observers had previously engaged in a social interaction. To determine whether social interaction primes the MR system, groups of participants engaged or did not engage in a social interaction before observing human or robotic actions. During observation, MR was assessed via motor-evoked potentials elicited with transcranial magnetic stimulation. Compared with participants who did not engage in a prior social interaction, participants who engaged in the social interaction showed a significant increase in MR for human actions. Social interaction did not increase MR for robot actions, indicating the potential role played by the history of interaction between an actor and an observer in action simulation. Thus, naturalistic social interaction and laboratory action observation tasks appear to involve common MR mechanisms, and recent experience tunes the system to particular agent types.

However, the computational principles underlying action interaction are still largely unknown. How can people judge the meaningfulness of interaction between two actors? Does the interaction between two dancers conducting the same dancing styles (e.g., two salsa dancers) look more meaningful than that between two dancers conducting different styles (e.g., one salsa dancer and one Indian dancer)? If so, is it due to the high level grouping effect of action style, or due to coherent alignment between limb movements at certain distinct moments, or just due to synchronization of dancing tempos? How much of this effect is contributed by top-down modulation (e.g., semantic knowledge, visual experience, or motor expertise)? Is the top-down

modulation also effective when interactive information is task irrelevant? All these questions would be worth addressing if we want to further investigate action understanding involving multiple actors, which are usually the case in most social situations.

1.3.2 Action prediction

A common feature of motion perception is "anticipation", which is one major component subserving precise event perception. For example, representational momentum (RM) effects (the offset position of a moving target is mislocalized in the direction of a movement), and boundary extension effects (the remembered view of a scene expands to include a region just outside the boundaries of the original view), are well-documented illustrations of the adaptive value of predictability in the visual system (see Thornton & Hubbard, 2002, for a review). More specifically, Bertamini (1993) found that memory for the position of an object is biased. When asked to judge whether an object has changed its position with respect to a position shown a few milliseconds earlier, observers tended to detect the displacement more often when the displacement was not in the expected direction (downward for a falling object). By measuring the memory distortion for the position of an object on an inclined plane, with quantitative changes of the inclination angle and retention interval (Figure 1-6), they found that the internal representation of an object is intrinsically dynamic and is influenced by the forces perceived as acting on the object, reflecting a natural tendency to mentally extrapolate the motion of a target stimulus into the future and our incapacity to stop the extrapolation of a moving target exactly at its disappearance (Finke, Freyd, & Shyi, 1986; Freyd, 1987). This extrapolation may involve two classes of cognitive operations, the cognitive motion extrapolation (CME) or the clocking/timing mechanism.

To investigate this problem, DeLucia and Liddell (1998) used the prediction motion (PM) paradigm in which an object moves toward a target and disappears; observers respond at the time that they think the object would reach the target. In previous studies (e.g., (Schiff & Detwiler, 1979; Schiff & Oldak, 1990), estimates of (time- to- contact) TTC increased as actual TTC increased, but observers estimated TTC inaccurately when actual TTC exceeded 2-3s, and the inaccuracy increased as actual TIC increased. For the two possible underlying mechanisms, cognitive motion extrapolation (CME) involves an internal cognitive representation of the object's visible motion. Rosenbaum (1975) proposed that observers may directly extend the

object's motion after it disappears, possibly relying on a form of imagery. Despite the large individual differences in eye movement patterns across experimental conditions or observers, the researcher still observed that, during the object's hidden motion in a PM task , participants most often moved their eyes continuously as if the target was still visible. In contrast, the clocking, or timing mechanism may simply compute the object's hidden time (i.e. time-to-contact, TTC) based on both the ratio of exposed distance to hidden distance and the duration of the object's visible motion (Lyon & Waag, 1995; Rosenbaum, 1975). Their results showed that the slope of the relationship between estimated and actual TTC in judgments of approaching objects decreased when visual information about the environment between the observer and the display was minimized. Moreover, the accuracy of relative duration judgments of visual (but not auditory) stimuli decreased when a PM task was performed concurrently, indicating that PM tasks involve cognitive motion extrapolation (tracking hypothesis) rather than solely a clocking process that counts down TTC (timing hypothesis).
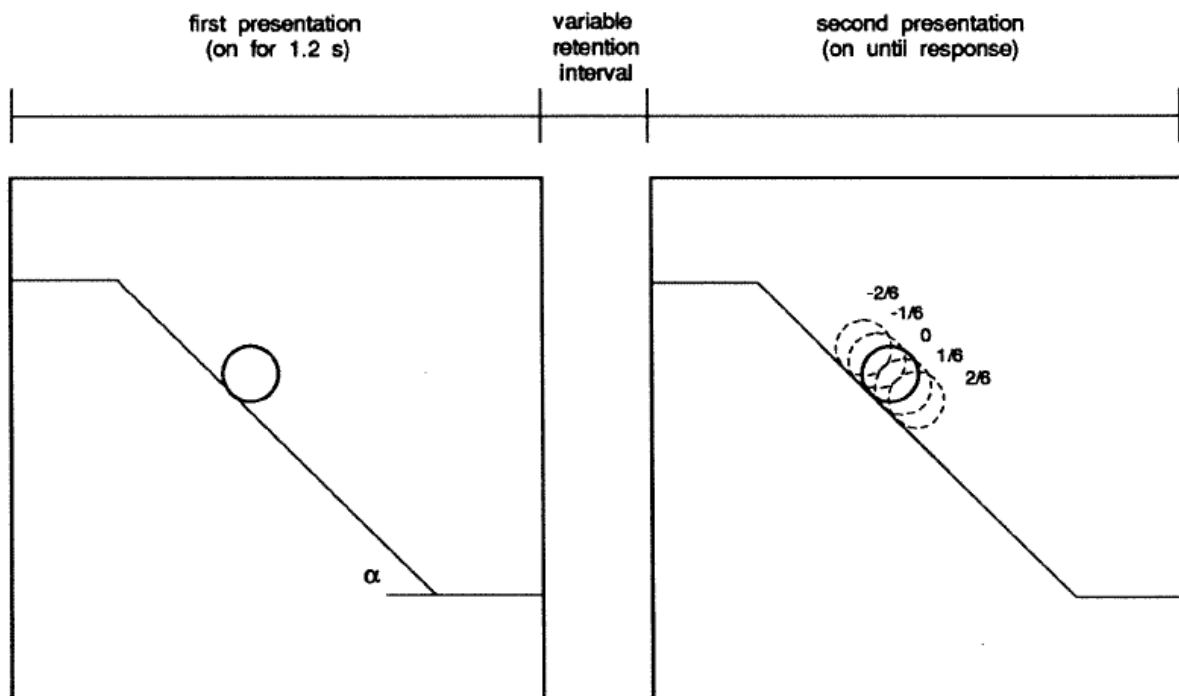
Figure 1-6. Description of the display. The angle α measures the inclination of the plane. The circle shown on the slope in the first frame disappeared during the retention interval. A probe circle reappeared in the second frame in one of five different positions. The numbers indicate the displacement in radii for each position.

As we mentioned above, action observation involves the participation of mirror neuron system, suggesting that action observation recruits the same cognitive process as motor execution. Thus, it would be natural to assume that action prediction involves action simulation, as if observers predict the next move of a person in the same way they imagine themselves to conduct the same action, which is termed "re-enactment" by Prinz (2006). He suggested that re-enactment exploits the motor system's inbuilt capacity to anticipate the consequences of action, and extends it from the planning of own action to the perception of foreign action. To test this theory, Graf and colleagues (2007) showed observers brief videos of point- light actions, followed by an occluder and a static test posture and requires asked them to judge whether the test stimulus depicted a continuation of the action in the same depth orientation. The occluder time and the movement gap (i.e., the time between the endpoint of the action and the test posture) were independently varied. Results showed that prediction performance was best when occluder time and movement gap corresponded, i.e., when the test posture was a continuation of the sequence that matched the occluder. This finding indicated that action prediction may involve a simulation process that operates in real-time. However, when actions were inverted, the matching effects disappeared, indicating the real-time simulation process broke down when the actions were presented under viewing conditions for which observers have little experience, either motor or visual, indicative of the role played by the mirror neural system.

Although the action-simulation process provides remarkable temporal precision in predicting actions, this simulation process may cost cognitive resources in the temporal domain producing consistent prediction errors. To test this hypothesis, Sparenberg and colleagues (2012) examined the time course of the real-time mental simulation process used to predict the outcome of occluded actions. In this study, participants were presented with transiently occluded point-light actions when manipulating the temporal outcome after occlusion, and asked to judge the temporal coherence of the action after a short and a long occlusion period. Results showed that observers revealed a consistent bias to judge the action during the occlusion as happening slower than the real time elapsed, indicating that action simulation took constantly longer than the observed action itself. Such a temporal judgment error in action prediction was absent when inverted actions were used, ruling out the hypothesis that such bias is a purely visually driven effect. Instead, the temporal error is probably due to costs arising from a switch from action perception to an internal simulation process involving motor representations.

More recently, it was found that this internal action simulation can also be modulated by action semantics. Springer and Prinz (2010) showed that action prediction accuracy can be influenced by a semantic effect. They found that compared to words that are action-unrelated (concrete nouns), processing of action verbs produces fewer errors when occluder time and movement gap match each other and more errors when they don't match (larger "diagonal distance effect"). This was also true when effects were compared between verbs describing dynamic actions and verbs describing static actions. Moreover, the degree of dynamics implied in action verbs or the speeding information (i.e., verbs expressing fast, moderate, or slow actions) (e.g., "to catch" vs. "to grasp" vs. "to stretch") can account for a slope variation of the diagonal distance effect. All these results indicated a linkage between action simulation and action semantics, a view that coincides with a recent notion of a close link between motor processes and the understanding of action language.

The involvement of action simulation during action perception may be specific for natural human movement. To investigate the influence of movement kinematics on the accuracy of action prediction, Stadler and colleagues (2012) created human point-light shapes either with human movement (natural condition), or with artificial movement that was more uniform regarding velocity profiles and trajectories (artificial condition). During brief occlusions, the participants were asked to judge after occlusion whether the actions were continued coherently in time or shifted to an earlier or later frame. They found that error rates and reaction times were increased in the artificial movement condition compared to the natural condition, indicating a perceptual advantage for movement with a human velocity profile, corresponding to the notion of a close interaction between observed and executed movement.

1.3.3 Prospective coding

The next natural question is: how is the computational model of action prediction implemented in human brains? What are the underlying neural mechanisms of that online real-time action simulation?

Although classical theories of sensory processing view the brain as a passive, stimulus-driven device, Engel and colleagues (2001) proposed that more research should examine the constructive nature of perception, viewing it as an active and highly selective process. More specifically, top-down factors can lead to states of 'expectancy' or 'anticipation' that can be

expressed in the temporal structure of activity patterns (i.e., synchronous oscillations) before the appearance of stimuli. Indeed, there is ample supporting evidence from both functional imaging and microelectrode studies on selective attention, working memory, or motor preparation. Those studies found that stimuli processing is controlled by top-down influences, which either increases baseline activities of neuronal firing or strongly shapes the intrinsic dynamic oscillation of thalamocortical networks and constantly creates predictions about forthcoming sensory events.

In studies investigating attention, delay matching and motion preparation, increased baseline activities biased toward expectation or anticipation were discovered, which provide top-down modulation. For monkey studies, during states of attentive expectation, baseline firing rates in monkey extrastriate areas V2 and V4 increase when attention is directed inside the receptive field of the recorded neurons (Luck, Chelazzi, Hillyard, & Desimone, 1997), reflecting the top-down signals that create a bias in favour of stimuli that will appear at the attended location. Functional imaging in humans has also provided evidence for stimulus-independent increases in activity during attentional expectation, (Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999; Rees, Backus, & Heeger, 2000). Similar conclusions can be drawn from studies of delay-period activity during working memory tasks, when sustained neural activity during delayed response tasks can encode, in a spatially selective and feature-selective manner, expected target stimuli (Luck et al., 1997; E. K. Miller, 2000). Neurons that produce such sustained activity have been found in many association areas, including prefrontal, posterior parietal and inferotemporal cortex, and also in subcortical structures (Hikosaka, Sakamoto, & Usui, 1989). In motor preparation, the activation of motor and premotor structures that precedes the execution of specific movements (Boussaoud, 2001; Crammond & Kalaska, 2000; Riehle & Requin, 1993) is thought to reflect specific processes, such as response selection, specification of movement parameters and the coordination of relevant neural populations. Activity in preparatory periods has been shown to predict, for instance, the direction, latency and speed of subsequent movements (Crammond & Kalaska, 2000; Dorris, Pare, & Munoz, 2000; Riehle & Requin, 1993), which may also be used to process action stimulation during action perception.

Further interesting questions would be: do these predictive top-down modulations also apply to action prediction? How would body configuration or global form modulate action prediction? How people predict local joint movement and global posture differently? How would action interaction affect action prediction? The answers to these questions are still largely

unknown and should be further investigated in order to fill the gap between computational models and neural mechanisms for action prediction.

1.4 Thesis overview

In this thesis, the main goal is to examine the perceptual or cognitive mechanisms underlying action interaction and action prediction, and their mutual influence. In Chapter 2, we examine whether interpersonal activities for human interactions (e.g., coupled salsa dancing), as opposed to isolated action of a single individual (e.g. walking), impact the visibility of stimuli in the very early level of visual processing. Using binocular rivalry paradigm, we systematically gauge what action information reaches awareness and what information does not. We find that actions eliciting natural body movements and involved in meaningful interpersonal activities are granted preferential access to visual awareness over actions that are not involved in interpersonal activities. In Chapter 3, we will investigate if social interactive cues are used in generating proactive predictions, by comparing them to non-interactive actions. Specifically, we intend to investigate whether action prediction processes differ between viewing two interactive actors and viewing a single actor. To examine different aspects of action prediction ('when' vs. 'what'), we asked participants to perform two different tasks. In the task emphasizing the 'when' information (i.e. predictive timing), observers were asked to judge the correct action continuation of the actor. In the second task emphasizing 'what' information (i.e. predictive identification), observers were asked to identify the testing actions after certain depth rotations. Our analysis specifically focuses on whether the presence of interpersonal actions influences the performance in the 'when' and 'where' tasks differentially.

In Chapter 4, we investigate how action prediction mechanisms interact with the generic mechanisms for a low-level visual task, localization of moving elements. To examine the mechanisms involved in encoding joint positions in a moving body, stationary dot was briefly flashed during the viewing of a walking action in a point-light display, and the magnitude of the flash lag (FL) effect was measured to assess the perceived location of moving joints. Given the dynamic nature of biological motion stimuli, we expected that participants would misperceive the relative position between flash joints and moving limbs in the action. A critical control condition with *inverted* actions was used to identify whether the FL effect depends on action processing. In addition, we examined how familiarity of an action, and the functional importance

of joints in an action, impact the FL effect in biological motion by comparing human performance with a baseline model. In this temporal averaging model (Krekelberg & Lappe, 2000), FL effect was elicited solely by generic mechanisms in motion processing.

In Chapter 5, we extended the *flash-lag* paradigm, to investigate the role of the global mechanisms that predict the posture change resulting from whole-body movements—other than the local mechanisms that predict future positions of joints involved in action prediction. To examine the global mechanism for predicting the whole-body posture in the near future, a stationary posture was briefly flashed during the viewing of a walking action. We expected that observers are more likely to judge that the walking posture lagged behind the flash, even though it is physically aligned with the flash. Taken together, our findings would provide psychophysical evidence that biological motion perception is predictive. By updating an internal model, people are able to anticipate the future position of critical joints and the future posture of whole-body movements.

Chapter 2 Social interactions receive priority to conscious perception

Humans are social animals, constantly engaged with other people. The importance of social thought and action is hard to overstate. However, is social information so important that it actually determines which stimuli are promoted to conscious experience and which stimuli are suppressed as invisible? To address this question, we used a binocular rivalry paradigm, in which the two eyes receive different action stimuli. In two experiments we measured the conscious percept of rival actions, and found that actions engaged in social interactions are granted preferential access to visual awareness over non-interactive actions. Lastly, an attentional task that presumably engaged the mentalizing system enhanced the priority assigned to social interactions in reaching conscious perception. We also found a positive correlation between human identification of interactive activity and the promotion of socially-relevant information to visual awareness. The present findings suggest that the visual system amplifies socially-relevant sensory information and actively promotes it to consciousness, thereby facilitating inferences about social interactions.

The human ability to deal with social stimuli enables us to recognize what others are doing and to understand why others act in certain ways, so that social interactions can be planned and executed. How the human visual system processes social stimuli is a fundamental question in social perception, as this mechanism is the basis for understanding of the goals and intentions of others (Allison, Puce, & McCarthy, 2000). Biological motion perception is a prime example of this type of processing (Rutherford & Kuhlmeier, 2013). Previous research has shown that sparse motion stimuli, consisting of just a few point-light dots representing joint movements of a human actor, can be readily analyzed to recognize detailed characteristics of the actor. These characteristics include gender, identity, action category, emotion and interpersonal interaction (Centelles, Assaiante, Nazarian, Anton, & Schmitz, 2011; Cutting & Kozlowski, 1977; Dittrich, Troscianko, Lea, & Morgan, 1996; Manera et al., 2011; Mather & Murdoch, 1994; Neri et al., 2006). Moreover, such simple displays with sparsely distributed moving dots suffice for inferences about the goals and beliefs of other actors (Manera et al., 2011; Van Overwalle & Baetens, 2009). Hence, the faculty of action perception and understanding entails a rather direct bridge between sensory representations and social cognition.

However, the complexity of the social world constantly challenges the capacity of the human visual system. Social scenes usually are cluttered with many actions involving different people and objects (e.g., imagine a scene in a busy train station or an airport). Deep processing of all actions and movements in the sensory input is difficult to achieve due to the limited capacity of the human visual and cognitive system. This inherent constraint may have entailed mandatory competition among observed actions for conscious experience. A central unresolved question in social cognition concerns how sensory information important for social inference is processed within the visual hierarchy (van Boxtel & Lu, 2015b). In the present study, we investigated whether the brain amplifies socially-relevant information and actively promotes it to consciousness, when observing human actions and inter-personal interactions, so as to facilitate social perception and inferences.

Previous research has demonstrated top-down influences of action perception on sensory processing. For example, recognizing meaningful actions biases depth estimation toward an interpretation consistent with human body structure, even when the physical depth is inconsistent with the layout of human body (Bulthoff, Bulthoff, & Sinha, 1998; Lu, Tjan, & Liu, 2006). Such findings suggest that knowledge about the structure and dynamics of human body movements guides the interpretation of sensory information in visual stimuli. Recent evidence also indicates that the presence of interpersonal interactions enhances sensitivity to detecting actions when the point-light stimuli are embedded in a noise background (Manera et al., 2011; Neri et al., 2006). This finding suggests that coordinated body movements between two actors impact on how actions of an individual actor are processed. These studies used action stimuli with relative short durations (range of 1 to 8 seconds). However, social interactions between agents in daily life extend for much longer periods. Over extended time, the availability of socially relevant stimuli to visual awareness may fluctuate, especially when other competing dynamic information in the visual scene requires access for processing. It remains unknown whether social actions and interactions receive priority to conscious perception compared to other non-social dynamic information.

There are reasons to anticipate that actions engaged in social interactions may receive priority to visual awareness. First, due to the importance of interactive activities in inferring social relations between agents, it would be beneficial for the organism to maintain higher sensitivity to such stimuli than to non-interactive actions. Indeed, previous research showed that

certain brain areas (in particular, the medial parietal and dorsomedial prefrontal cortices) are jointly recruited to yield increased activity when observing social interactions (Iacoboni et al., 2004). Second, it has previously been shown in binocular rivalry displays (i.e. displays in which the two eye's inputs compete for subjective awareness) that contextual cues play important roles in determining which stimulus reaches conscious awareness. For example, when two gratings are presented in the two eyes, each eye's view is dominant for approximately equal amounts of time. However, when one of the two stimuli is embedded in a larger display, this global context causes it to be more visible (D. Alais & Blake, 1999; Sobel & Blake, 2002). These studies showed that contextual information biases the selection processes of sensory input at an early stage of visual analysis. This process may also apply to social stimuli. Indeed, in social situations involving two interacting people, a rich set of contextual cues becomes available, originating from coordinated movements and a common goal. In these situations, the actions of one actor dictate, to a high degree, the actions performed by the companion actor. These contextual cues may play an important role in selecting and promoting action information to conscious perception.

Therefore, we adapted a binocular rivalry paradigm to examine the impact of social action stimuli on some of the earliest stages of visual processing. When the two eyes receive markedly dissimilar patterns, observers experience the phenomenon of binocular rivalry: at any moment in time one of the patterns is perceived as the dominant stimulus, but perception alternates between the rival patterns every few seconds (David Alais & Blake, 2005). These fluctuations afford a sensitive measure of the impact of stimulus variables (e.g., contrast) or object-based properties on rivalry dynamics (Levelt, 1968), thereby shedding light on how the brain constructs our visual perceptions.

When two action stimuli, each consisting of a dozen disconnected point-lights, are presented to different eyes and put into conflict in a binocular rivalry paradigm, one might expect that these disconnected dots would undergo rivalry independently of each other. However, a previous study showed that the point-lights representing a walking action tended to undergo rivalry as a grouped entity (Watson, Pearson, & Clifford, 2004), suggesting that the visual system uses knowledge about human body movements as a way to structure the input.

Here we examine whether social interactive activities, as opposed to solitary actions from a single individual (e.g. walking), impact the rivalry between two competing action stimuli to

determine what action information reaches awareness and what information does not. In the present study, salsa dancers were used to generate rivalry stimuli, taking advantage of the rich whole-body movements in the social action sequence and the high coordination of movements between the two dancers. We expected that actions eliciting natural body movements that are involved in meaningful interpersonal activities would be granted preferential access to visual awareness over actions that are not involved in interpersonal activities, when measuring visual dominance in binocular rivalry displays.

2.1 General methods

The study was approved by IRB #12-000277-CR-00004. Consent forms and debriefing documentations were provided. Stimuli were created using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) and were displayed on a calibrated Viewsonic CRT monitor with a resolution of 1280 × 1024 pixels. Participants viewed the stimuli through an adjustable stereoscope from Berezin Stereo Photography Products in a constant viewing distance of 57 cm maintained by a chin rest. All actions were obtained from the Carnegie Mellon Graphics Lab Motion Capture Database (http://mocap.cs.cmu.edu). The BioMotion Toolbox (van Boxtel & Lu, 2013b) was used to convert the raw motion capture files to the point-light display.

On each trial two rival point-light actions were presented, one to each eye and in different colors (i.e., blue and red) to induce binocular rivalry. The rival actions were displayed in the same retinal region of the two eyes for a relatively long duration in the range of 20 to 26 seconds. Actors subtended 7.3 by 5.3 degrees of visual angle. Point-lights were shown with luminance level of 10.6 cd/m^2 for all colors (red, green or blue) on a black background (~0 cd/m^2). The size of each point-light was 0.13°. A central fixation cross (with size of 0.75° by 0.75°) was presented, as well as a frame surrounding the stimuli on the screen to help subjects register the same position in the two eyes. The size of the frame was 13.28×13.28° with a line width of 0.075°. In order for the rival actions to spatially overlap as much as possible to create sufficient rivalry perception, we eliminated extrinsic motion of the body by fixing the mid-point of the two hip joints of the rival actors at the same position across the entire trial.

Previous research (Blake, 1989; Blake, Yu, Lokey, & Norman, 1998) has shown that motion signals with speed around 1.2 deg/s strongly attract dominance in rivalry, yielding exclusive visibility of one eye's view (Blake, 1977). To rule out the possibility that the perceived

speeds of smoothly moving joints determine the rivalry dynamics, we randomly sampled point-lights along the limbs, with a limited lifetime of one frame using the method developed by Beintema and Lappe (2002). Previous research showed that humans still can readily recognize actions presented in the limited-lifetime display. This method removes the smooth motion trajectories of displayed dots, thereby eliminating potential contributions from inter-frame local motion signals in binocular rivalry, allowing us to focus on effects attributable to action processing based on posture change over time. As shown in Figure 2-1, point-lights were randomly sampled along the limbs and were displayed with a limited lifetime of one frame (i.e., 13ms). Each limited lifetime actor was composed of nine dots, including the head and eight dots randomly sampled on each of the 8 limb segments.
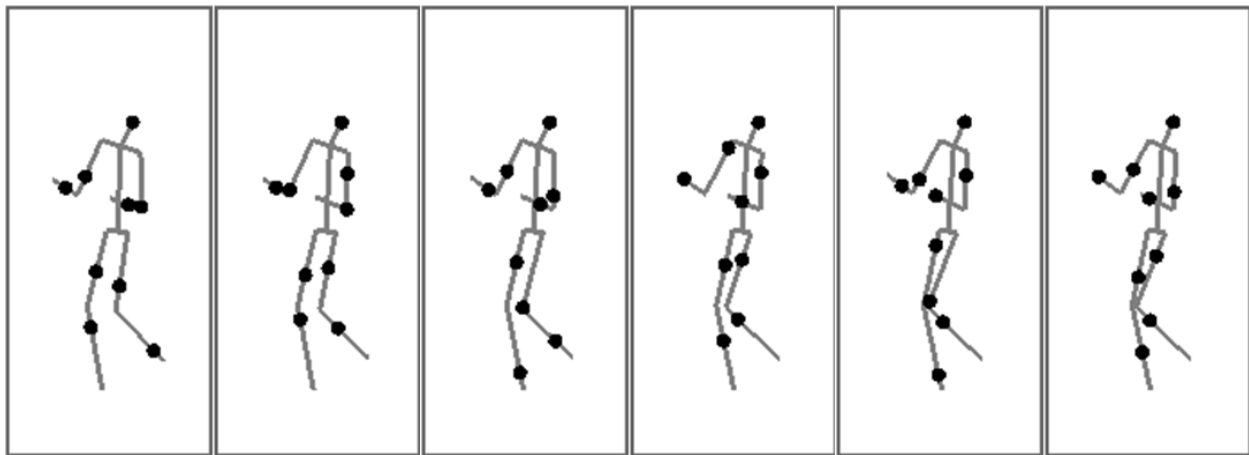
Figure 2-1. An illustration of stimulus frames presented in the limited-lifetime display. Dots were randomly sampled along the limbs and were displayed for one frame. The lines are shown just for the purpose of demonstration.

For all the experiments reported in the present paper, participants were asked to press and hold one of three keys to indicate whether the "blue" (left arrow) or "red" (right arrow) dots were more visible, or the two groups of colored dots were equally visible (down arrow), at any moment throughout the trial.

We excluded participants who failed to satisfy two criteria: (a) over all trials combined, an observer pressed a specific response button (corresponding to red or blue rival action) more than 95% or less than 5% of the time; (b) there were more than 16 trials (corresponding to 50% trials in Experiment 1) in which an observer only pressed the button corresponding to "mixed" percepts during the entire trial. These criteria applied to all experiments.

2.2 Experiment 1: Natural activities are more visually dominant than inverted activities

Previous research reported that two upright point-light walkers, each with a different facing direction, elicited stronger rivalry than two inverted walkers (Watson et al., 2004). This finding suggests that actions with more ecologically-relevant upright orientation are granted a preferential access to visual awareness. This rivalry result is consistent with the well-documented inversion effect in the literature on biological motion perception (Sumi, 1984). However, when an upright action is placed in direct conflict with an inverted action, it is still unknown whether the upright action will gain visual dominance and/or the inverted action will be suppressed. Experiment 1 aims to address this question by measuring visual dominance when displaying an upright action to one eye, and an inverted action to the other eye. This experiment examined whether the rivalry inversion effect of a familiar action (e.g. walking, (Watson et al., 2004)) can be generalized to relatively unfamiliar actions (e.g., dancing), and whether this effect is still observed when limited lifetime dots (i.e. each point change positions randomly on the limb across frames) removes inter-frame motion trajectory information. We predicted that the upright dancer would receive preferential access to awareness (i.e. stronger predominance in rivalry) when competing with an inverted dancer.

2.3.1. Methods

Participants completed a practice session using the walking rival stimuli as in Watson, Pearson and Clifford. (Watson et al., 2004). The practice session included four trials showing two upright walkers with different facing directions (leftward or rightward) presented one to each eye, in different colors (red or blue), and four extra trials showing inverted walkers as the rival stimuli, in intermixed order. Participants were asked to press and hold a button to indicate the color of the dominant dots (red, blue or mixed) throughout the 26-second trials.

In each trial of the subsequent test session, one of eight female salsa dancers was selected to generate both rival actions, one upright and one inverted (see Figure 2-2A). The stimuli were presented dichoptically: an upright dancer to one eye and an inverted dancer to the other eye, each in a different color. The rival actions were displayed with a different random sampling of the limited lifetime point-lights. An inverted dancer is often still recognizable as human activity but with much lower recognition accuracy. For example, Dittrich (Dittrich, 1993) found that

recognition accuracy for inverted waltz dancer was 61% in a free response task. Hence, inverted dances serve as a social but less ecological stimulus in rivalry. Participants were asked to indicate the color of the dominant actor (red, blue or mixed) continuously throughout the entire trial. The rival stimuli lasted for 26 seconds on each trial. The experiment consisted of 32 trials for each of the upright and inverted conditions.

2.3.1.1 Participants

All participants were undergraduate students at the University of California, Los Angeles (UCLA) and participated for course credit. All observers had normal or corrected-to-normal visual acuity. Sixteen observers (13 females, average age of 20.7) participated in Experiment 1. One participant's data were excluded from the analysis based on the above described exclusion criteria. The sample size was estimated according to relevant studies on binocular rivalry with social stimuli in the literature (Watson et al., 2004) and pilot studies. The number of around 20 subjects is consistent with the sample size used in binocular rivalry studies if using naïve subjects. The exact number of participants was determined by a stopping rule of convenience, in that we stopped collecting data at the end of the week in which we achieved as least 20 participants (this was less in experiment 1, based on the cited literature).
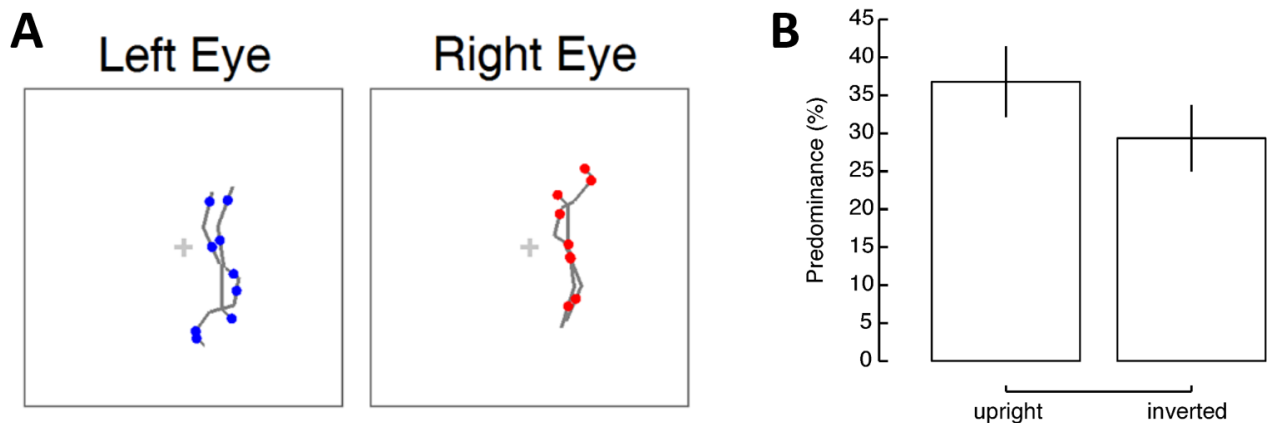
Figure 2-2. Stimulus illustration and results in Experiment 1. (A) Schematic illustration of rival actions in different colors in dichoptic presentation. One eye viewed an upright dancer and the other eye viewed an inverted dancer. The gray lines are only for illustration purpose to show the randomly sampled dots along the limbs, and were not shown in the experiment. (B) Results of Experiment 1 are presented as predominance, for both upright and inverted dancers. Error bars indicate the 95% confidence interval throughout the paper.

2.3.2 Results

The predominance (the percentage of reported dominance over the total viewing duration) for upright dancers (M = 36.81%) was significantly greater than that of the inverted dancers (M = 29.35%; t(14)=2.43, p = .029, Cohen's d = .83; see Figure 2-2B). We also examined a second measure of binocular rivalry, the average time duration of visual dominance for each rival stimulus. We found that upright dancers were visible for longer durations than inverted dancers (upright, M = 4.49 s; inverted, M = 4.13 s; t(14)=2.30, p = .037), providing converging evidence that actions with the more ecologically-relevant upright orientation receive precedence to visual awareness compared to inverted actions.

2.3.3 Discussion

The present results bolster and extend the findings reported by Watson et al.(Watson et al., 2004) (2004). Our results suggest that when upright and inverted actions directly compete with each other in a rivalry setup, the visual system employs the ecologically-relevant upright body orientation as a reference to group the visual input into meaningful and coherent units, even for dancing actions that observers do not regularly perceive or perform in daily life. The present experiment employed the limited-lifetime technique to alleviate the potential contribution of local motion mechanisms, such as a "life-detector" based on characteristic movements of the feet in recognizing and detecting walking actions (Troje & Westhoff, 2006; van Boxtel & Lu, 2015a). Hence, the present paradigm allows us to identify rivalry effects primarily attributable to global action processing.

2.3 Experiment 2: Interactive activities are more visually dominant

Experiment 2 aimed to examine whether an action engaged in social interaction receive precedence in access to consciousness compared to a solitary action performed by a single actor. Binocular rivalry was created by presenting a male salsa dancer to one eye, and a non-salsa actor to the other eye. We predicted that without showing the interacting partner, the two rival actions (salsa dancer and non-salsa actor) would show comparable visual dominance during rivalry. On the contrary, when presenting a binocularly-viewed actor performing a dance with one of the rival actors (i.e., in the presence of social interaction), this social salsa dancer would receive preferential access to visual awareness compared a non-interactive actor.

2.3.1 Methods

In Experiment 2, the rival dichoptic stimuli were presented in a limited-lifetime display, one in red and the other in blue (counterbalanced across trials). The rival stimuli consisted of a male salsa dancer, randomly selected from one of the four salsa couples, and a non-salsa actor (an exuberant laugher or an Indian dancer; see Figure 2-3A for an illustration). The rival non-salsa actions included most characteristic movements of the salsa dance, such as arm waving, leg lifting, and body shaking. We ensured that low-level stimulus characteristics, such as the average inter-frame speed of joint movements, and the size of the actors, were matched between the rival salsa dancers and the competing non-salsa actors. Figure 2-4 depicts the summary of the matched average inter-frame speeds between the rival actions used in the experiment. We thus minimized the potential difference of low-level visual features in the two rivalry stimuli during binocular rivalry. In the *one-actor* condition, a single actor was presented to each eye including the male salsa dancer and the non-salsa actor to provide the rivalry actions. In the *two-actor* condition, the stimuli included the same rivalry actions as those in the one-actor condition but also with a third actor presented to both eyes in the left side of the visual field. This binocularly-viewed non-rival action was generated from a female salsa dancer that was partnered with the male salsa dancer at the time when the actions were recorded, thus displaying a truly inter-personal activity. As shown in Fig 3A panel, the binocularly-viewed non-rival action was shown as a green stick figure with the line width of 0.038 visual degrees to minimize the effort needed to recognize this partner dancer. The rivalry stimuli lasted for 20 seconds in each trial. The experiment consisted of 64 trials. As in Experiment 1, participants were asked to indicate the color of the dominant actor (red, blue or mixed) at any moment throughout the entire trial.
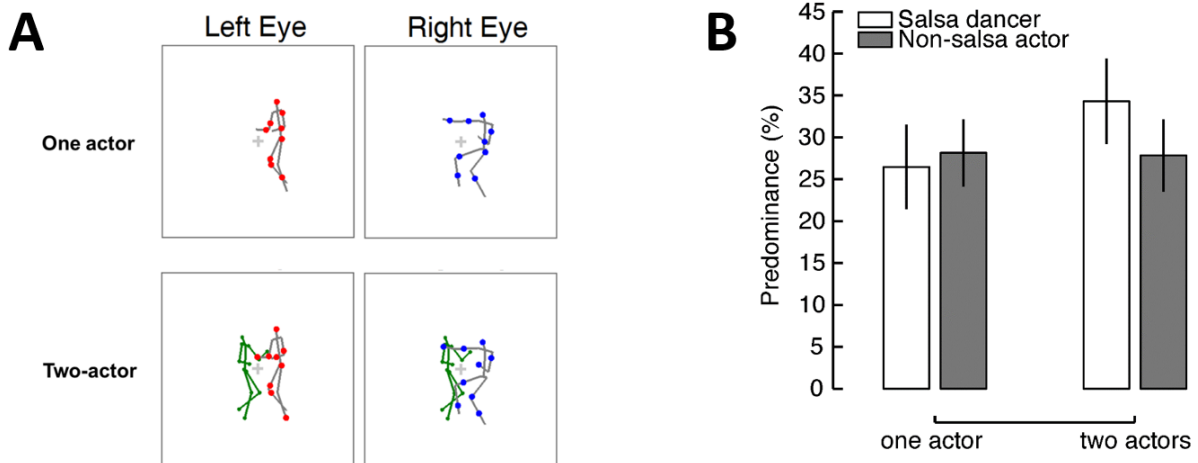
Figure 2-3. Stimulus illustration and results in Experiment 2. (A) Schematic illustration of a frame from the rival stimuli in dichoptic presentation. The partner of the rival salsa dancer was binocularly presented in green in the two-actor condition, and was absent in the one-actor trials. The two rival actions were shown in a limited-lifetime display (the gray lines are only for illustration purpose). In this illustrated stimulus at the bottom, the two actors shown to the right eye did not engage in a meaningful interaction, whereas the two actors presented to the left eye performed a salsa dance (a meaningful interaction between two actors). (B) Results of Experiment 2. The predominance difference between the two rival actions was significantly greater in the presence of the partnered dancer than in its absence.
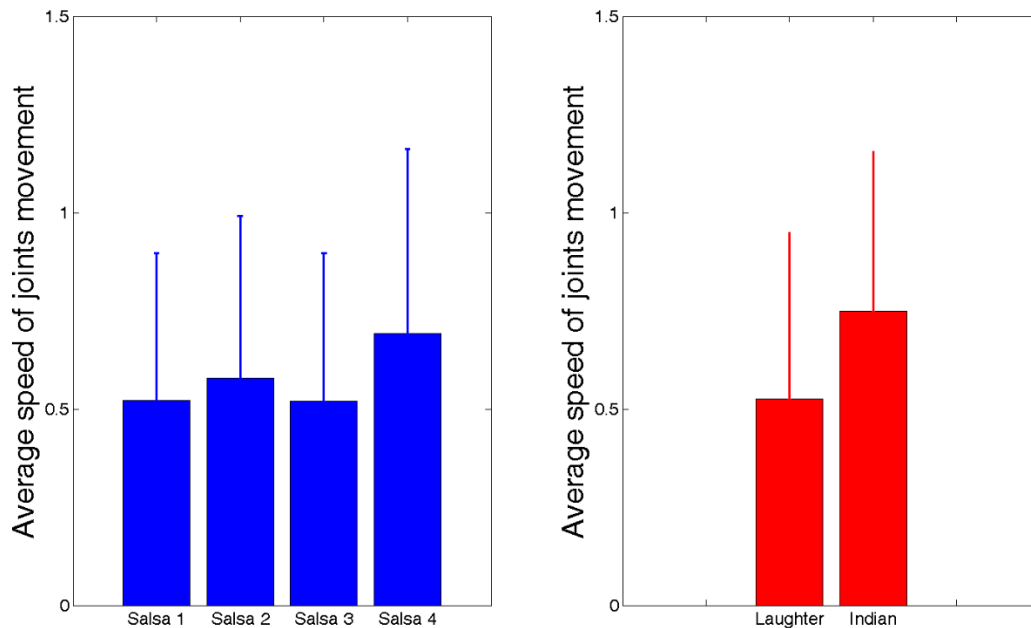
Figure 2-4. The average inter-frame speed (pixel/frame) of joint movements for rival actions used in the Experiments 2 and 3. Error bars represent standard deviations.

2.3.1.1 Participants

26 observers (22 females, average age of 20.6) participated in Experiment. No observers participated in more than one experiment in the present paper. According to the exclusion criteria, five participant's data were excluded in the analysis.

2.3.2 Results

As shown in the Figure 2-3B, when the binocularly-viewed dance partner was absent, the two rival actions did not differ in visual predominance (male salsa dancer, $M = 26.46\%$; non-salsa actors, $M = 28.15\%$, $t(20) = 0.99$, $p > .250$), indicating that there was no inherent visual dominance preference for the salsa dancer over the non-salsa actor. The equal visibility of the two rival actions indicates that matching of low-level stimulus characteristics between the two actions during binocular rivalry was successful.

However, in the presence of a binocularly-viewed partner dancer, the rival salsa dancer became more dominant, receiving precedence to visual awareness compared to the rival actor

who did not engage in a social interaction. This result is supported by a significant two-way statistical interaction effect ($F(1,20)=8.34$, $p = .009$, $\eta_p^2= .29$) in a repeated measures ANOVA on predominance with two within-subject factors, rival action types (salsa dancer/non-salsa action) and interactivity (one-actor/two-actor) . We also found that the significant increase in predominance of the salsa dancer when the partnered dancer was displayed compared to when the partnered dancer was not displayed (presence, $M = 34.30\%$; absence, $M =26.46\%$ $t(20) = 3.60$, $p = .002$, Cohen's $d = .66$).

Additional analyses on the average duration of visual dominance, provided converging evidence; the difference of the dominance durations between the two rival actors depended on the presence of the partnered dancer ($F(1,20)=10.33$, $p = .004$, $\eta_p^2= .34$). Specifically, when the partnered salsa dancer was displayed (despite being viewed binocularly), the dominance duration for the rival salsa dancer significantly increased compared to that when the partnered dancer was absent (5.54s with partner vs. 4.33s without partner, $t(20) = 2.69$, $p = .014$, Cohen's $d = .33$). The consistent findings from both predominance and dominance duration suggest the existence of a mechanism that boosts priority for promoting interactive actions to visual awareness. This mechanism does not act on non-interactive actions, as there was no evidence of increased suppression for non-interactive actions.

2.4 Experiment 3: Inversion reduces the impact of interactive activities on visual dominance

It might be argued that the rivalry behavior in Experiment 2 is driven by certain mid-level visual features, rather than social interpersonal activity. Such mid-level features include symmetry or movement coordination between the partnered salsa dancers. These critical features could serve as informative cues to trigger the mechanisms of grouping the two dancers, and consequently render the interactive action more visible. This impact could conceivably take place without the identification of the social content contained in action stimuli. In order to address this issue, Experiment 3 was designed to include two within-subject sessions, one showing upright actions and the other with inverted actions. The inversion manipulation significantly weakens effective identification of action and interactive activity (Dittrich et al., 1996; Neri et al., 2006; Sumi, 1984), but maintains mid-level features such as symmetric

postures and coordinated dot movements between partnered dancers, and other potential grouping cues in the stimuli with interactive activity.

2.4.1 Methods

Experiment 3 included one session identical to Experiment 2, and the other session in which the actors were inverted in all trials. The order of the two sessions was counter-balanced across participants. The inversion manipulation reduces holistic action processing (Dittrich et al., 1996; Neri et al., 2006; Sumi, 1984), which consequently weakens the percept of interactive activity, while maintaining motion cues and inter-personal coordination identical to that in the upright condition.

2.4.1.1 Participants

Twenty-six observers (17 females, average age of 21.2) for Experiment 3. All participants satisfied the inclusion criteria and were included in the analysis.

2.4.2 Results

First, Experiment 3 replicated the findings of Experiment 2 for the predominance data for the upright actions, indicated by a significant two-way interaction effect between the interactivity (one-actor/two-actor) and rival action type (salsa dancer/non-salsa action) ($F(1,25)=11.37$, $p = .002$, $\eta_p^2 = .31$). This finding confirms that the predominance difference between the rival actions with ecological body orientation (i.e. upright display) depended on the presence of interactive activity in the display.

When the actions were inverted, the impact of interactive activity on the predominance difference was reduced to trend level ($F(1,25)=4.18$, $p = .052$, $\eta_p^2 = .0.14$). This marginally significant result appears to suggest that coordinated movements between two actions may play a role in determining the predominance measure even if the actions are displayed upside-down. However, we found an influence of the block order (inverted session first or upright session first) on predominance, revealed by a significant four-way interaction effect ($F(1,24) = 6.055$, $p = .021$) in a mixed ANOVA analysis with the order of the blocks as a between-subjects factor, and orientation (upright/inverted), interactivity (one-actor /two-actor) and rival action type (salsa dancer/non-salsa action) as within subjects factors.

Inversion only caused the social interaction effect to disappear when the inverted session was run first (Figure 2-5). This result suggests that the influence of inversion was susceptible to a learning effect when observers familiarized themselves with coordinated salsa dancing in a prior block with the upright display. This learning effect is consistent with previous findings that humans appear to be able to "learn" to see inverted action. For example, Hiris and his colleagues found that people can learn to detect the presence of inverted biological motion nearly as well as they detected upright actions (Hiris, Krebeck, Edmonds, & Stout, 2005). When focusing only on the participants who performed the inverted block first, and hence lacked the learning opportunity of improving recognition with inverted actions, we found that the impact of interactive activity on the predominance difference was significant in the upright condition ($F(1,13) = 9.28$, $p=0.009$, $\eta_p^2 = 0.42$), but not in the inverted condition ($F(1,13) = 0.99$, $p = 0.34$, $\eta_p^2 = 0.07$). These results suggest that, without prior exposure to point-light displays, coordinated movements and symmetry cues preserved in the inverted condition were not sufficient to elicit a significant influence on visual dominance of interactive actions.
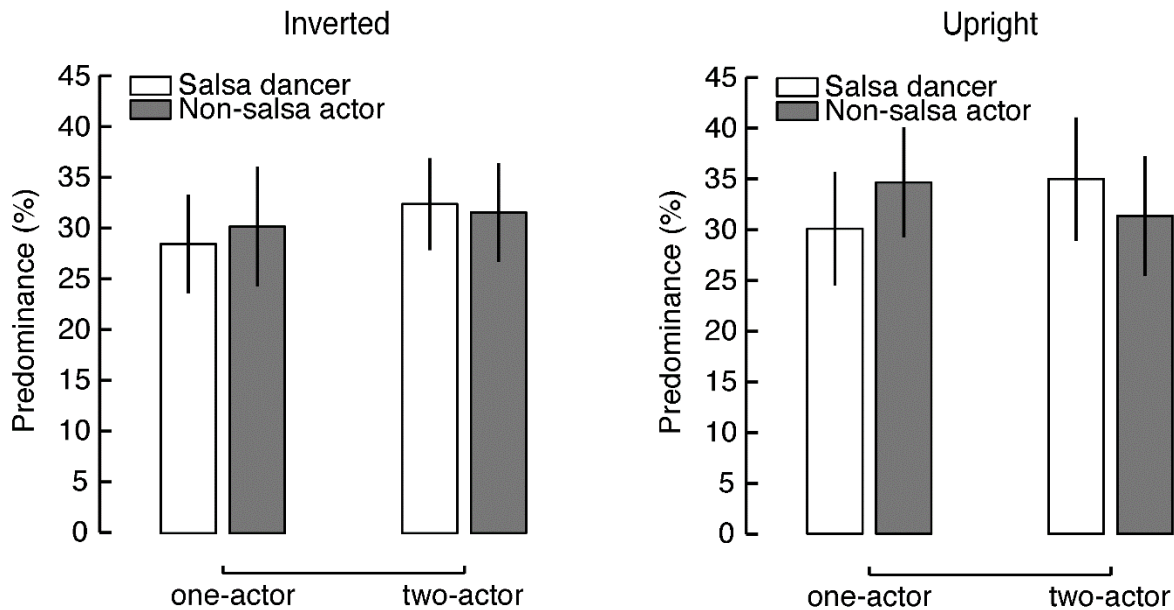
Figure 2-5. Results of Experiment 3. Predominance values for the participants who performed the inverted block (left) before the upright block (right).

Similarly, we found that the difference of the dominance durations in upright rivalry depended on the presence of the partnered dancer, as indicated by a significant two-way interaction effect ($F(1,25)=7.94, p = .009$). In contrast, the same effect was not significant ($F(1,25) = 2.72, p = .111$) in the inverted condition. These results suggest the impact of interactivity on maintaining visual dominance was reduced when rival actions were presented upside-down, likely due to the weakened perception of inverted actions.

2.5 Discussion for Experiment 2 and 3

The findings in Experiment 2 from two binocular rivalry measurements (i.e. percentage of total dominance and average duration of each dominance) show that social interactions presented in the upright body orientation propel relevant actions into conscious awareness. Results in Experiment 3 further suggest that inverted displays of the actions reduced such advantage for the social interactions to enter conscious awareness.

Hence, the gating of social actions into awareness may be most effective when interactive activity is apparent or easy to extract from visual inputs with ecological body orientations (i.e., upright). In the visual world, social interactions usually involve two agents coordinating their body movements. Coordinated motion can be a strong predictor of interactive activity. Hence, consistent with findings supporting the critical role of movement coordination in joint action (Knoblich, Butterfill, & Sebanz, 2011; Obhi & Sebanz, 2011), perception of coordinated body movements might enable the promotion of potential socially-relevant information to visual awareness to a certain degree, even when social interaction appears difficult to perceive in some situations (e.g., the inverted display in our paradigm, when viewed after a block of upright displays).

2.6 Experiment 4: Identification of interactive actions promotes visual dominance

Identifying the presence of an interactive activity with upright actions was quite effortless in Experiment 2 and 3, because the partnered salsa dancers were either absent or present. It has been suggested that when actions are simple and goal-directed, the analysis can be performed by an automatic process (Frith & Frith, 1999). However, more complicated tasks, such as

identifying inter-personal interaction and inferring social intentions, require attention (Bargh & Williams, 2006; de Lange, Spronk, Willems, Toni, & Bekkering, 2008; Keysers & Gazzola, 2007; Spunt & Lieberman, 2013). In Experiment 4 we increased the difficulty of the tasks by introducing the third binocularly-viewed salsa dancer in all conditions (i.e. partnered vs. un-partnered). We also engaged attention by explicitly asking participants to identify whether two agents were performing meaningful interactive activities besides performing the primary task of tracking the visual rivalry. We predicted that partnered dancers would receive more visual dominance in rivalry compared to un-partnered dancers when attention is explicitly demanded in discriminating genuine inter-personal interactions from two actors performing similar actions independently without coordination.

2.6.1 Methods

Experiment 4 used the same rival stimuli as in Experiment 2: a male salsa dancer presented to one eye, and a non-salsa actor (either a laughing person or an actor performing an Indian dance) to the other eye. A third actor was shown to both eyes as a green stick figure. In the *partnered* trials, the third actor was the female partner dancer engaged in a salsa dance with the rival male dancer so as to form a dynamically-coupled interpersonal interaction. In the *un-partnered* trials, the third actor was a male salsa dancer from one of the other recorded dance couples, thus sharing a similar action style and statistical regularity of kinematics with the rival salsa dancer, but not engaging in interactive dancing with the rival male actor. An alternative way of removing the synchronicity between agents (which indicates interactivity) is to split the action sequence into two sequences and cross-pair the sequence of each individual actor, as was done in a previous study (Neri et al., 2006). We did not use this manipulation because some of the recorded dancing actions were not long enough (i.e. longer than 40 seconds) to allow minimal trial durations of 20 seconds. Figure 2-6 illustrates the rival stimuli and the trajectories in the two conditions of this experiment.
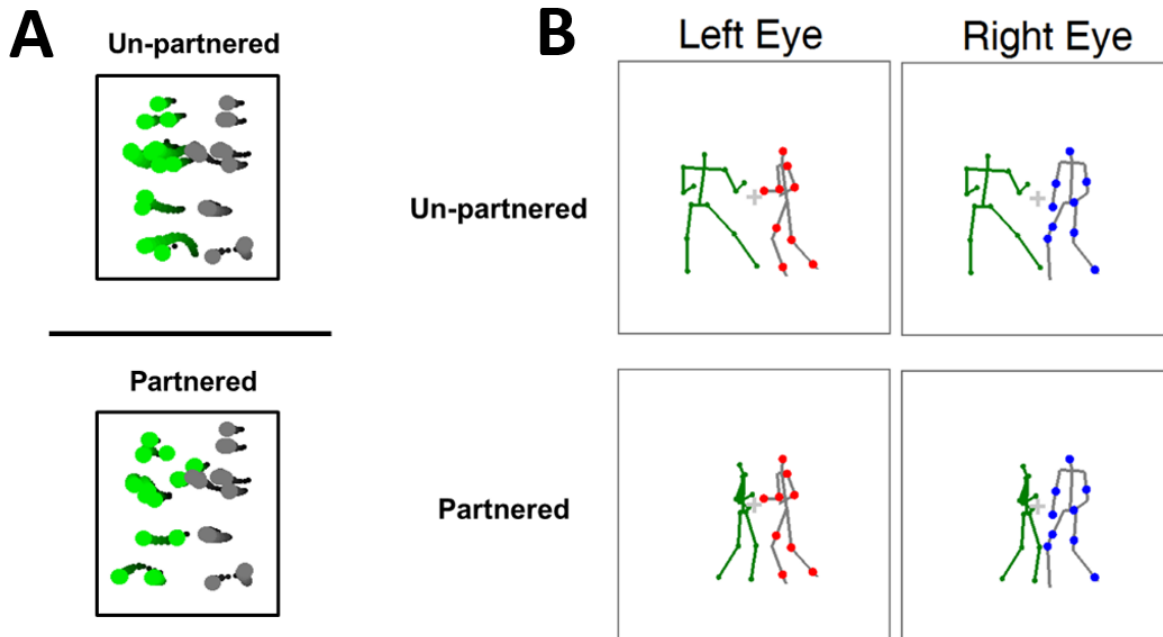
Figure 2-6. Schematic illustration of stimuli used in Experiment 4. (A) The trajectories of joint movements for a pair of un-partnered two male salsa dancers and a coupled salsa pairs. To give an impression of movement, this schematic shows several frames with increasing dots sizes for more recent frames. In this example, the two dancers in the un-partnered condition did not engage in a coordinated activity, whereas the two dancers in the partnered condition performed a salsa dance (a meaningful interaction between two actors). (B) The stimuli included two rival actions displayed in red and blue, and a binocularly viewed action displayed in green.

Participants were asked to indicate the color of dominant dots during binocular rivalry, with the same instructions as in the first two experiments. In addition, at the end of each trial, they reported whether any two actors in the display had performed interactive activities. This second task was designed to engage attention to social interactions between agents in the observed actions. Experiment 3 consisted of 64 trials, each lasting for 20 seconds.

2.6.1.1 Participants

Twenty-three observers participated in Experiment 4. One participant's data were excluded from the analysis based on the exclusion criteria.

2.6.2 Results

With respect to identifying the presence of a meaningful interaction, participants achieved a modest performance level with mean accuracy of 0.69 (SD = 0.15), significantly better than chance ($t(21)$ =5.77 , p < .001), though considerably less than perfect. The modest level of performance in distinguishing partnered salsa dance couples from un-partnered dancer pairs is consistent with findings in a recent study (Thurman & Lu, 2014). Hence, the identification of interpersonal interactions in the present experiment was not trivial, but rather required effort and focused attention.

In this experiment, we found that attentional focus on identifying meaningful social interactions significantly enhanced the visibility of the rival salsa dancer. As depicted in Figure 2-7A, the predominance of the rival salsa dancer increased when the partnered dancer was present compared to its rival non-salsa actor (salsa dancer, M = 38.21%; non-salsa actor, M = 28.20%, $t(21)$ = 4.38, p < .001, Cohen's d = .94). In contrast, the same comparison in the un-partnered condition did not yield any difference (salsa dancer, M = 31.34%; non-salsa actors, M = 30.00%, $t(22)$ = 1.39, p = 0.180). The difference in predominance between the two rival actors thus depended on whether the third actor is was a partnered dancer or an un-partnered dancer performing similar dance movements ($F(1,21)$ = 11.11, p = .003, η_p^2 = .35). We found that the predominance of the rival salsa dancer significantly increased in the presence of a meaningful interaction (partnered, 38.17%; un-partnered, 31.31%; $t(21)$ = 3.03, p = .006, Cohen's d = .61). Consistently, longer dominance durations were also obtained for the rival salsa dancer with interaction than without (partnered, 4.32 s; un-partnered, 3.69 s; $t(21)$ = 2.22, p =.038, Cohen's d = .19). As for the non-salsa action, predominance for was reduced slightly with interaction compared to without (partnered, 28.18%; un-partnered, 29.98%; $t(21)$ = 2.55, p =.019, Cohen's d = .54). However, this was not true for dominance durations (partnered, 3.82 s; un-partnered, 3.63 s; $t(21)$ = 0.61, p =.546).
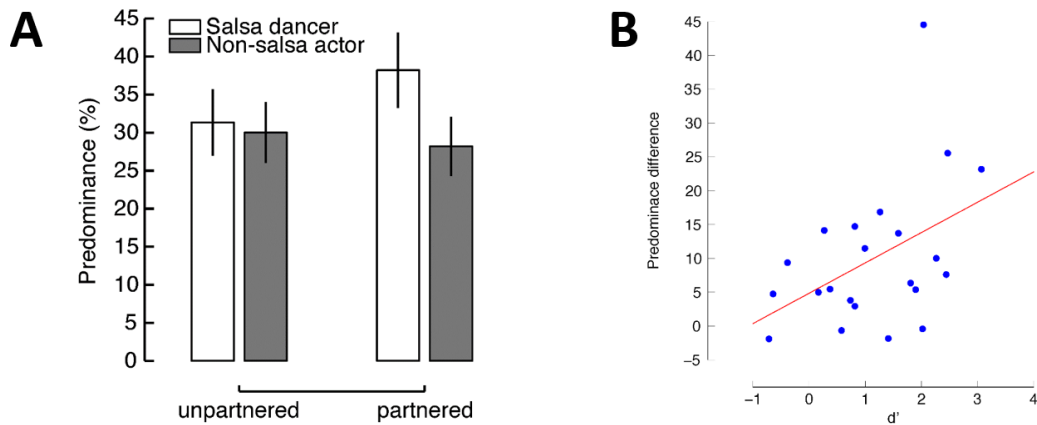
Figure 2-7. Results of Experiment 4: interactivity promotes visual dominance. (A) The predominance difference between the two rival actions was significantly greater in the presence of a partnered dancer than of an un-partnered dancer. Error bars indicate the 95% confidence intervals. (B) Scatter plot showing the relationship between predominance difference between the two rival actions in the partnered condition and sensitivity (d') in identifying interactive activities.

We further analyzed the relationship between predominance in the rivalry task and sensitivity in the identification task. The predominance difference between the two rival actions (i.e. the rival male salsa dancer and the non-salsa actor) in the partnered condition significantly correlated with sensitivity (d') in identifying interactive activities ($r = .44$, $p = .040$, see Figure 2-7B; removing the potential outlier did not affect these data $r = .45$, $p = .043$). This finding suggests a positive association between ability in identifying social interactions and selectivity in promoting socially-relevant information to visual awareness. We also examined the potential relation between rivalry performance and response bias. We found that the subjective bias favoring the response of "interactive" did not correlate with enhanced visibility of interactive actors in the rivalry task ($r = -0.12$, $p = .610$). In addition, un-partnered trials which were misidentified as containing interactive activity showed no difference between the two rival actions in predominance ($t(21) = 1.51$, $p = 0.147$) or mean dominance duration ($t(21) = 0.64$, $p = 0.528$), excluding the possibility that participants merely reported the salsa dancer as dominant as long as they reported interactivity. Both results suggest that a response bias is unlikely to account for our main findings.

It might be argued that people have a bias to report or perceptually force the rival salsa dancers to be more visible because of its similarity with the binocularly-viewed female dancer

(as both belong to the same action category, and were also task-relevant during the entire experiment). If this were the case, a similar effect would be expected in the un-partnered condition, since the binocularly-viewed dancer was a male salsa dancer selected from a different dancing couple thus showing even higher similarity with the rival male salsa dancer. However, we found equal predominance for the two rival actions in the un-partnered condition, ruling out this type of response bias as the cause of our finding.

2.7 General Discussion

In the present study we found that actions eliciting natural body movements and meaningful interpersonal activities are granted preferential access to visual awareness. We employed binocular rivalry between natural unfamiliar activities, that contained a social interactions or not. Our approach contrasts with previous studies on rivalry of socially-relevant stimuli in the following important way: We investigated how social content in terms of *inter-personal interactions* influences the formation of visual awareness for human actions, whereas previous studies used non-interactive social stimuli (faces or isolated actors).

Previous studies that have used faces as social stimuli often employed continuous flash suppression (CFS) paradigm (Tsuchiya & Koch, 2005), a potent masking paradigm similar to binocular rivalry. It has been found that upright faces break into visual awareness more quickly than inverted faces (Jiang, Costello, & He, 2007), and some emotional faces break through especially quickly (Yang, Zald, & Blake, 2007). However, it has been argued that such effects may be due to low-level feature differences between rival stimuli (Coelho, Cloete, & Wallis, 2010), rather than the facial aspects per se (but see (Stein, Senju, Peelen, & Sterzer, 2011)).

Crucially, in the present study we controlled for low-level stimulus differences, and ensured that the rival actions were equally dominant without social interactions. These rival actions with equal stimulus strength made it possible to scrutinize how social interactions influence which action stimuli receive priority to conscious perception, and which action stimuli are suppressed as invisible. For rival actions that were equally visible in isolation (i.e., without the social context), we found that the inclusion of social interactions for one of rival actions significantly enhanced visibility of that action.

Our results on social interaction are reminiscent of previous studies showing contextual influences in binocular rivalry for simple grating stimuli (e.g., (D. Alais & Blake, 1999)). In that

study, surrounding grating stimuli influenced the dominance balance of a pair of rivalry gratings. Analogously, the social contextual cues in interpersonal interactions impacted the rivalry dynamics in our stimuli. Furthermore, we found that interactive activities only increased the dominance durations of the action eliciting the inter-personal interactions, and did not affect the dominance duration of the non-interacting actor. This finding is consistent with what is known from binocular rivalry with grating stimuli (Sobel & Blake, 2002), where it has been shown that the dominance duration of gratings increased if the rival grating was consistent with the context, while the dominance duration of the inconsistent stimulus remained unchanged. Experiments on similar simple stimuli also showed that attention only impinges on the dominance of the attended stimulus, and does not affect (i.e. suppress) the dominance of the non-attended stimulus (Chong, Tadin, & Blake, 2005). We speculate that the same mechanisms are at work in our stimuli, with both contextual social information and attention being capable of increasing the dominance of the context-congruent stimulus, but leaving the context-incongruent stimulus largely unaffected.

Understanding actions, and social interactions in particular, requires an elaborate analysis of the visual input, including the context in which the actions take place. There is increasing evidence for a dual-processing strategy for understanding social actions and interactions. According to this theoretical framework, the first system analyzes social stimuli in an automatic manner, without the need for attentional focus, mainly through the mirror neuron system (Bargh & Williams, 2006; Spunt & Lieberman, 2013; van Boxtel & Lu, 2012). Even in the task of person recognition, a recent study (Rice, Phillips, Natu, An, & O'Toole, 2013) showed that humans automatically make use of body information to identify other individuals. This "automatic" system may rely on an embodied or "mirrored" simulation of the other's actions to interpret observed activities or understand the underlying goals (de Lange et al., 2008; Keysers & Gazzola, 2007). These types of actions may include simple locomotion, such as walking, or simple goal-directed actions, such as reaching. The second system supports controlled social causal attribution, and is voluntarily engaged when a deeper understanding of actions is required (Sperduti, Guionnet, Fossati, & Nadel, 2014). The function of this so-called mentalizing system (Frith & Frith, 2006) can be impaired by high cognitive load or enhanced by attentional focus (Spunt & Lieberman, 2013). Interestingly, fMRI studies (Centelles et al., 2011; Ciaramidaro, Becchio, Colle, Bara, & Walter, 2014) have shown that brain networks involved in the mirror and mentalizing systems are concurrently active when observing social interactions.

The findings in the present study are consistent with and further support the dual-process account. In the first three experiments, we showed that meaningful social (inter)actions gain preferential access to consciousness without requiring focused attention to social interactions (i.e. when the interaction is task-irrelevant). This effect is plausibly due to an automatic action-analyzing mechanism (presumably the mirror-neuron system). In Experiment 4 we showed that performing an attentional task specifically increased the access of social actions to consciousness. In a pilot experiment without the second task, we did not find the impact of interactivity (with partnered dancer vs. un-partnered dancer) on visual dominance of rival salsa dancer, suggesting that the attentional task is required. More critically, individual differences in the ability to identify interactive activity in complex scenes were correlated with increases in access of social actions to consciousness. It has previously been shown that attention impinges on the mentalizing system and not (or less so) on the mirror-neuron system (Spunt & Lieberman, 2013), suggesting that the increase in the visual consciousness of interactive actors in Experiment 4 was likely due to an increased engagement of the mentalizing system. These findings extend previous research on individual differences in action perception (Kaiser & Shiffrar, 2009; L. E. Miller & Saygin, 2013; van Boxtel & Lu, 2013b) by suggesting an explicit link to the mentalizing system. By comparing the results in the last three experiments, we conclude that the involvement of automatic vs. attentional effects will depend on the specific comparison conditions in the experiment. If the comparison involves comparing the presence versus the absence of a cooperative dancer, the impact of social interactions is likely initiated automatically, because this is a relatively easy comparison (c.f. Experiment 2 and 3). However, when the comparison involves comparing subtle differences in interactivity (e.g., partnered dance couple vs. two actors performing similar dance movements; Experiment 4), attention engagement to social interaction is critical to actively modulate the visual dominance of the relevant action.

Binocular rivalry suppression is often thought to occur through interoccular inhibitory interactions between monocular channels (Blake, 1989; Brascamp, Sohn, Lee, & Blake, 2013) that determine which stimulus reaches awareness (the dominant stimulus) and which stimulus does not (the suppressed stimulus). Our results show that social cues have a major impact on which visual information is gated to visual awareness, by exerting a large influence on interocular suppression, probably via cortical feedback (Blake & Logothetis, 2002; van Boxtel,

Alais, & van Ee, 2008). Our findings therefore imply that in controlling what stimuli can reach awareness, the human brain gives precedence to subtle socially relevant stimuli over non-social stimuli, very likely through contributions of both the mirror-neuron system and the mentalizing system. This selectivity to social cues enables exquisite sensitivity to socially relevant information, even before the visual stimuli reach full consciousness, especially when the visual inputs are noisy and ambiguous and competing for awareness or attention resources. Social cues are thus in the vanguard of information reaching conscious awareness, and that is made available for purposeful executive functions such as planning and contemplation (Crick & Koch, 1995), thereby enabling us to function effectively and efficiently in the complex social world in which we live.

Chapter 3 Meaningful interaction and action prediction

Vision plays an important role in perceiving other people's actions, moods and intentions. Although static pictures convey rich information, dynamic motion pictures generally provides a better temporal precision and more meaningful interpretation (Blake & Shiffrar, 2007). Humans, as highly social creatures, rely heavily on the ability to recognize others' motion sequences and predict what the next posture would be. However, different circumstances may require a different emphasis in the type of prediction. For example, when a single action is viewed, and it generally does not take much effort to interpret the action, instead, the most important thing to the brain is to correctly predict the timing of the action continuation (predictive timing). However, in more complex situations in which people interact, it may be more fruitful to predict what is going to happen instead of exactly when (predictive identification).

Human point-light actions (PL animations) are excellent candidate stimuli to investigate human motion perception in the aspect that they provide only dynamic cues of the body movements in the absence of all other complicated visual information such as form, texture, color or brightness. We use PL actions in our study to ensure that we measure people's action prediction solely based on motion signal. Researches actually showed that human visual system is highly sensitive to sparse visual stimuli of biological motion when there is a single actor. Even by reducing the actor to a set of 11 moving dots and presentation time to less than one-tenth of a second still allow for meaningful interpretation (Jansson & Johansso.G, 1973).Also, partial occlusion such as viewing through apertures also allows for recognition (Lu, 2010; Shiffrar, Lichtey, & Chatterjee, 1997). Furthermore, humans can still recognize point-light actions embedded in a noisy background (Bertenthal & Pinto, 1994; Cutting, Moore, & Morrison, 1988; Neri, Morrone, & Burr, 1998; Thompson, Hansen, Hess, & Troje, 2007), assigned with random contrasts (Ahlstrom, Blake, & Ahlstrom, 1997) or jittered positions (Beintema & Lappe, 2002), associated with scrambled depth (Bulthoff et al., 1998; Lu et al., 2006), or assigned with a general transformation such as inversion (Bertenthal & Pinto, 1994; Bertenthal et al., 1984; Pavlova & Sokolov, 2000; Pinto & Shiffrar, 1999; Simion et al., 2008; Sumi, 1984) and even when spatially scrambled (Chang & Troje, 2009b; Troje & Westhoff, 2006). The exquisite sensitivity to biological motion are demonstrated by humans from a very young age (Blake & Shiffrar, 2007) since infants four months old will stare at human motion sequences for longer

durations than they will at the same number of dots undergoing random motions (Fox & Mcdaniel, 1982). People not only detect and recognize actions from sparse biological motion stimuli, they can also extract sufficient information to infer future actions in real time based on their prior knowledge of the temporal regularity in daily human motions.

Given the sparse motion signals, people can do action prediction with substantial temporal resolution of several hundred miliseconds (Graf et al., 2007; Jarraya, Amorim, & Bardy, 2005). Also, brief motion sequences of human actions produced a larger priming effect (shorter reaction time) when test static postures are the correct continuation of the motion sequences rather than when the test postures proceeded the motion sequences or are unrelated to the motion sequences (Verfaillie & Daems, 2002). Elecrophysiological studies in macaques also indicated that neurons in superior temporal sulcus (STS) respond to the sight of a static posture that followed a particular body action (e.g. articulated motion), but not when if followed other actions (Jellema & Perrett, 2003). In contrast, the effective action or posture presented in isolation or in different sequences failed to produce a vigorous response. In fact, even a static image of an action by itself induced activation in motion responsive areas (Kourtzi & Kanwisher, 2000), (suggesting that the brain interprets visual stimuli in a pro-active manner (Bar, 2007). Although, single actor prediction possesses considerable temporal precision, action processing may be different when there are two interactive actors with more social cues.

Visual comprehension of human actions promotes effective social interaction. Humans exhibit fine discrimination ability in perceiving socially relevant characteristics from highly degraded depictions of human actions such as identity (Cutting & Kozlowski, 1977; Loula, Prasad, Harber, & Shiffrar, 2005; Troje et al., 2005), gender (Barclay, Cutting, & Kozlowski, 1978; Kozlowski & Cutting, 1977; Mather & Murdoch, 1994; Sumi, 2000; Troje & Geyer, 2002), emotion (Atkinson, Dittrich, Gemmell, & Young, 2004; Clarke et al., 2005; Dittrich et al., 1996; Pollick, Paterson, Bruderlin, & Sanford, 2001; Roether, Omlor, Christensen, & Giese, 2009; Walk & Homan, 1984), sexual orientation (Ambady, Hallahan, & Conner, 1999), dancing ability (Brown et al., 2005), openness (Brownlow, Dixon, Egbert, & Radcliffe, 1997), social dominance (Montepare & Zebrowitzmcarthur, 1988), vulnerability to attack (Gunns, Johnston, & Hudson, 2002) and even intent to deceive (Runeson & Frykholm, 1983).

Since action estimation can improve people's cognitive or motor function in social life, vice versa, social signals can help action perception in many ways. To function better in daily life, people need to response immediately and accurately to other people's actions at each moment. Examples include dancing a waltz, playing a piano duet, carrying a table together when actors direct their attention to where the interaction partner is attending (joint attention) (Manera et al., 2011). Most relevant to our study is the research showing that social context aids action perception. For example, observers detect the presence of angry PL walkers more accurately than neutral, happy, sad or fearful PL walkers (Chouchourelou, Matsuka, Harber, & Shiffrar, 2006). Also, meaningful interactions facilitated the visual discrimination of actors in noisy background (Neri et al., 2006). In both cases, researchers have found empirical evidence that communicative actions or meaningful interactions can enhance visual discrimination or visual detection of human agents in biological motion. It should very interesting to investigate how social interactive cues can influence action prediction.

In this study we will investigate if social interactive cues are used in generating proactive predictions, by comparing them to non-social situations. These different circumstances may require a different emphasis in the type of prediction. More specifically, we intend to investigate whether the prediction of others' action changes when people view two interactive actors rather than a single actor. To address different aspects of action prediction ('when' vs. 'what'), we used two tasks to address this question. One is a prediction task, in which observers are asked to perform 2-alternative forced choice (2AFC) and choose, among the two static postures, the correct action continuation of the actor in the proceeding probe movie that we made invisible for various amounts of time. This task emphasizes the 'when' information (i.e. predictive timing) In the other task we asked the subjects whether the testing dynamic stimulus is rotated in depth or not compared to the actor in the probe movie. This task emphasized the 'what' information (i.e. predictive identification). We specifically focus on whether social information influences the 'when' and 'where' tasks differentially.

3.1 Experiment 1: Choosing between two static figures

3.1.1 Methods

3.1.1.1 Participants

Twenty-one subjects participated in this experiment (2 male, 19 female; mean age 20±1.7 y). Subjects were given course credit for their participation, and all were naïve to the purpose of the experiment. All procedures received prior approval from the UCLA IRB.

3.1.1.2 Stimuli

We obtained the action stimuli from the motion-capture database created by Carnegie Mellon University online (http://mocap.cs.cmu.edu). It provides the three-dimensional coordinates for all 13 joints of the actors (the two feet, knees, hips, wrists, elbows, shoulder joints, and the head) of each time point. The 3D positions of the joints were projected on the screen using orthogonal projection. The height of figures was 6.8º. Point-light displays were created by displaying colored dots and lines (green, black, red or blue with dot diameter of 0.53º) on a white background (Figure 3-1 shows example stimuli). The initial rotation angle was selected for each movie sequence so that when there are two actors interacting with each other, they are seen from the side view (Figure 3-1 shows an example trial).

We used seven two-actor interactive movie sequences: 'A pulls B, B resists'; 'A pulls B, B resists, but loses'; 'A and B scramble for last seat'; 'Scramble for last seat, A loses and stands up'; 'High-five' (Figure 3-1 shows example trial of this movie sequence); 'A sits; B pulls up A'; and 'A picks up stool, threatens to strike B'. Among the two actors, actor A is the critical actor that we asked subjects make action prediction about without ignoring actor B if B is there. To familiarize subjects with both the type of stimuli and the task, we used another 5 one actor movie sequences in a practice session: 'Alternate squats'; 'Bend, lift'; 'push heavy object'; 'get up from chair'; and 'throwing'.
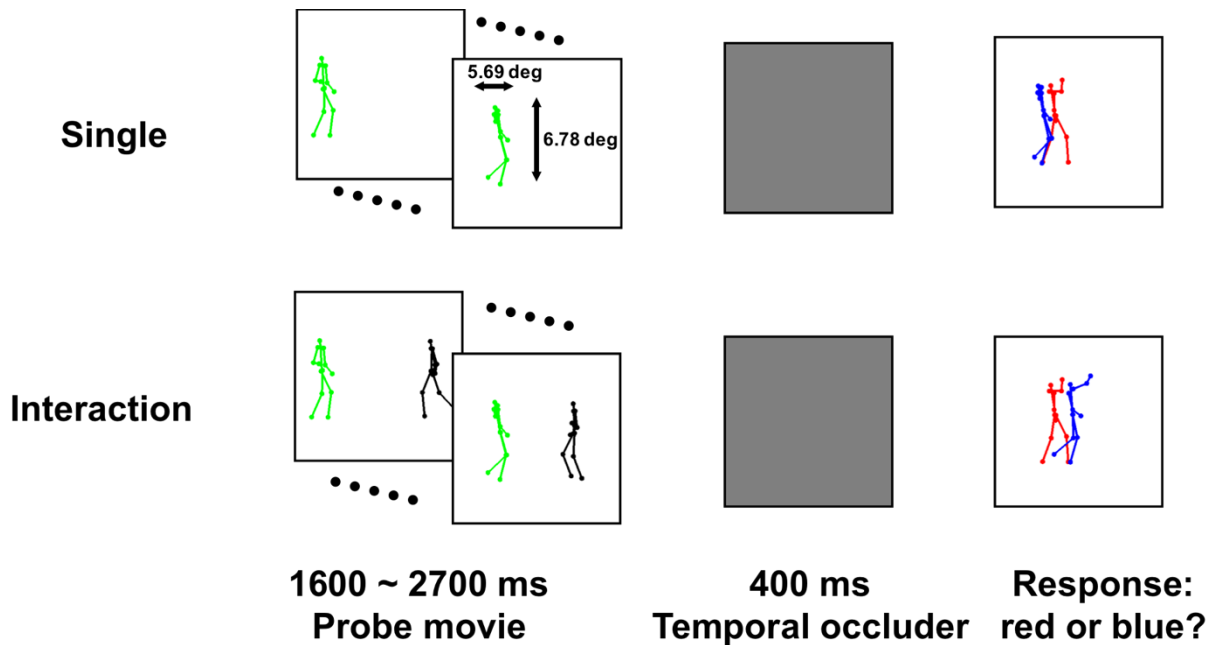
Figure 3-1. A sample trial for either single condition (upper) or interactive condition (lower) when the movie sequence is 'High-five'. The dynamic probe movie lasts for 1600~2700ms for all the seven movie sequences (1667ms for 'High-five'). The temporal occlude lasts for 400ms and the static test stimuli last for 253ms.

3.1.2.3 Procedure

Subjects used a chin-rest to maintain a distance of 47 cm from the screen. Screen resolution was 1208 pix × 1024 pix and the refresh rate 75Hz.

Each trial started with a 1.33 second pause during which on-screen instructions informed observers the completed / total number of trials. Then we presented a probe movie lasting about 1600~2700ms (different length for different movie sequences) on a white background screen. In the probe movie, there can be either a single actor (green) or two interactive actors (green and black). The green human is always the actor A, the critical actor which we asked subjects to do action prediction about. At the end of the probe movie, the actions were interrupted with a 400ms temporal occluder (blank gray screen of average luminance); we let subjects image the lights are 'switched off' and figures are still moving). After that, we presented two static test poses that were spatially overlapping, one being red and the other being blue. One of these poses was the correct continuation/pose of the green figure in the probe movie (target) after the blank, and the other was a pose (distractor) extracted from a different time point of the green figure from the

same movie sequence with a movement gap other than 400ms (it could be any of the eight: 400ms, -200ms, 0ms, 200ms, 600ms, 800ms, 1000ms, and 1200ms as shown in Figure 3-2). The colors (red or blue) of the test stimuli were counterbalanced for target and distractor from trial to trial. The movement gap is defined as the time elapsed between the end of the probe movie and the frame selected for the static test frame of the distractor. When the movement gap is negative, that indicated that the distractor is extracted from the probe movie before the end of the probe movie. The 2 test poses (one is target and the other is distractor) will be displayed on screen until subjects make the response. They have to pick out the target by indicating the color of the target (red or blue?). Sound feedback (2 beeps, a higher pitch following a lower pitch) was given to subjects when their answer was correct. Only correctness was analyzed to demonstrate subjects' action prediction performance.

To summarize, there were eight different movement gaps of distractors (-400ms, -200ms, 0ms, 200ms, 600ms, 800ms, 1000ms, and 1200ms) and two kinds of probe movies (Single vs. Interactive). Among all the sixteen conditions in the formal trials, each condition was repeated 28 times (7 movies × 2 colors × 2 viewpoints (left side view or right side view)), yielding a total of 448 trials. All trials were displayed in random order.

Before the experiment, observers completed a practice session of 20 trials. The practice session was nearly identical to formal experiment when we presented two test stimuli to let subject perform prediction task after the probe movie and the blank screen. The only difference was that we presented single actor probe movies (mentioned in the Stimuli part above) to make the display as simple as possible. This served as a preparation for subjects to get familiar with the task. The task again is to indicate which is the correct action continuation (red or blue?).

3.1.2 Results

The results show that subjects generally performed the task well (i.e. above 75% when the trials are easier, Figure 3-2) A repeated measure ANOVA revealed a main effect of Movement Gaps ($F_{(1, 20)}=17.250$, $p<0.0005$), a non-significant main effect of the number of actors we showed in the probe movie (single- vs. two-actor) and an interaction between Movement Gaps and single versus interacting actors ($F_{(7,140)}=3.566$, $p=0.022$).

The U-shaped dependence of accuracy on movement gap indicated that the more similar the distractor was to the target, the more difficult it was for subjects to tell the correct action continuation. However, the significant interaction in the ANOVA indicated that when the distractor had occurred in the preceding movie sequence (probe movie), observers were more accurate in predicting future action when viewing interactive action than single action. On the other hand, when target and distractor were both sampled from future postures (especially after the temporal occluder), subjects were more accurate in predicting action with a single actor. Subsequent paired t-tests suggested significant differences between the single and interactive conditions when the movement gap is -200ms, 0ms, and 600ms (reversed effect compared to -200ms and 0ms). The effect is 5.1% (t(20)=2.251, p=0.036) for -200ms, 7.5% (t(20)=2.695, p=0.014) for 0ms, and -5.3% (t(20)=2.223, p=0.038) for 600ms.
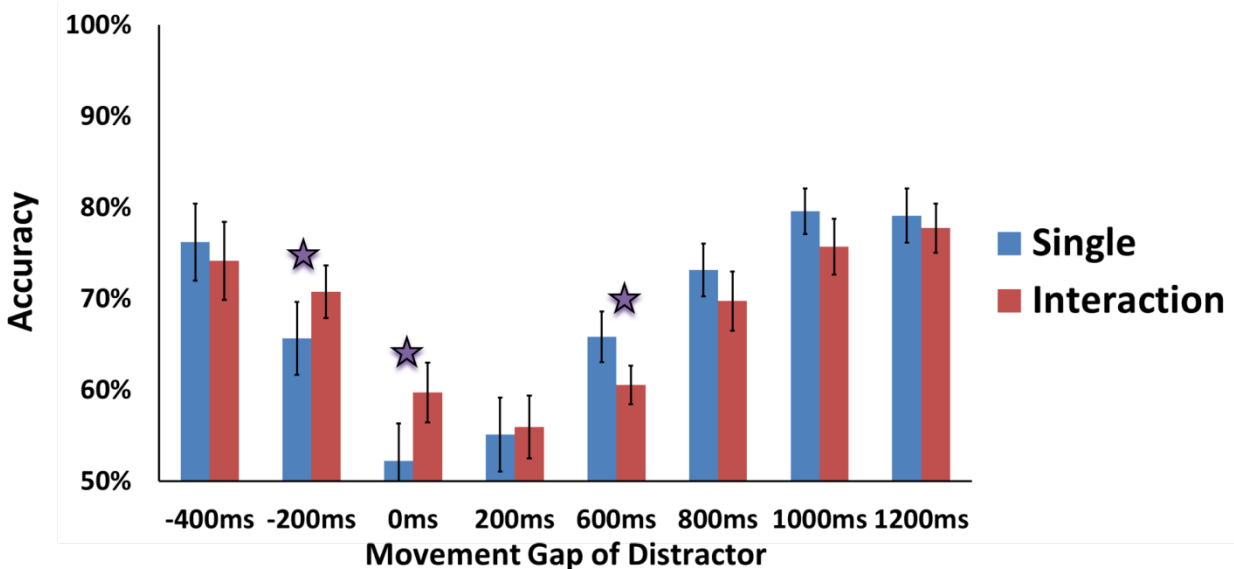
Figure 3-2. Bar graph of accuracy rates of all the sixteen conditions. Blue bars represent single conditions. Red bars represent interactive conditions. The x-axis indicates eight movement gaps (-400ms, -200ms, 0ms, 200ms, 600ms, 800ms, 1000ms, and 1200ms). Stars above bars indicated significant paired t tests between single and interactive conditions within any movement gap condition.

3.1.3 Discussion

We were particularly interested in determining whether people can predict what the upcoming postures should be after certain temporal occlude. More importantly, we wanted to know how their prediction can be influenced if there are two interactive actors compared to when

there is a single actor. This experiment provided evidence for real-time simulation in the prediction of PL actions. Even when the action is occluded, internal models seem to run time-locked to ongoing inferred actions. When the difference between the movement gaps of the target and the distractor becomes larger, it is easier to discriminate the incorrect posture, either earlier or later than what the actor should be, from the correct match posture. This indicates that observers were able to form a specific prediction about what the posture should be at this specific time point (400ms after the probe movie).

More interestingly, when the distractor had occurred in the preceding movie sequence (especially when movement gap is -200ms or 0ms), observers were more accurate in predicting future action when observing interactive action than single action, suggesting that observers may excel in action sequence processing, which means they are better in the 'what' dimension (i.e. predictive identification). However, when the target and the distractor were both sampled from future postures (especially when movement gap is 600ms), subjects were more accurate in predicting action with a single actor, suggesting a greater temporal precision in the single-actor prediction ('when' or predictive timing).

Difference between single-actor condition and interpersonal condition cannot be attributed to sole attention effects because otherwise the overall performance difference should be consistent across various movement gaps. The dissociation of predictive accuracy indicated that one-actor or two-actor motion sequences exerted influences on separate aspects of action prediction.

3.2 Experiment 2: Judging the rotation or not in depth

Instead of directly asking subjects to predict the correct posture after temporal interruption, which may tackle the 'when' problem more than 'what', we want to measure whether a second interactive actor will help people's action prediction in the 'what' dimension when we let subjects to do a rotation detection task. Since rotation detection is addressing the change of action posture and motion trajectory, our hypothesis is that the rotation detection task is presumable more sensitive to "what" information than "when" information, which should improve performance relevant to interaction conditions over single-actor conditions.

3.2.1 Methods

3.2.1.1 Participants

Fifteen subjects participated in this experiment (6 males, 9 females; mean age 20.3±5.8 y). Subjects were given course credit for their participation, and all were naïve to the purpose of the experiment. All procedures received prior approval from the UCLA IRB.

3.2.1.2 Stimuli

Stimuli were identical to those in Experiment 1. However, in the end of the trial, the 3D figures were either rotated 45 degrees in depth (i.e. vertical to the screen surface) or not.

3.2.1.3 Procedure

Similar to Experiment 1 except for a few changes indicated below. We presented a probe movie lasting about 520~1090ms (different length for different movie sequences and shorter than Experiment 1 because the test stimuli are dynamic now). In the probe movie, there can be either a single actor (red) or two interactive actors (red and blue). Their actions were not finished and continued during the temporal occluder (blank screen of either 400ms or 700ms) when the whole screen became gray (we let subjects image the lights are 'switched off' and figures are still moving). After that, we presented one dynamic figure (red), which was sampled from the action trajectory of the red figure in the probe movie with movement gap of either 400ms or 700ms, and either rotated 45º (inward or outward of the screen in depth) or not. Movement gap also referred to the time between last frame of probe movie and first frame of dynamic test in the correspondent movie sequence. When the duration of the temporal occluder equals to that of the movement gap, the trial is called a 'congruent' trial (movement gap is 400ms when occluder is 400ms, or movement gap is 700ms when occluder is 700ms); when the duration of the temporal occluder doesn't equal to that of the movement gap, the trial is called an 'incongruent' trial (movement gap is 400ms when occluder is 700ms, or movement gap is 700ms when occluder is 400ms).The test stimulus was presented for 253ms and then disappeared. Subjects were asked to give the response at this time (Figure 3-3 for sample trials). Subjects were asked to indicate whether the dynamic test stimulus was rotated or not after the blank screen by pressing one of the two buttons on the keyboard (left arrow key indicating the test stimulus was 'rotated'; right

one indicating 'not rotated'). Sound feedback (2 beeps) was given to subjects when their answer was correct.

Among all the sixteen conditions, each condition was repeated 14 times (7 movies × 2 viewpoint (left side view or right side view)), yielding a total of 224 trials. All trials were displayed in random order.

Before the experiment, subjects completed a 24-trial practice sessions which was identical to main experiment. The only difference was that we presented other single actor probe movies (mentioned in the Stimuli part above) to make the display as simple as possible. This served as a preparation for subjects to do the rotation detection task in the formal experimental trials. Furthermore, we added a familiarization session after practice session to show the seven whole original movie sequences to subjects twice, with text caption on top of the screen explaining the meaning of the movies (mentioned in Stimuli above). Subjects did not need to do any task, just passively viewing all the 7 two-actor movie sequences that will be used in the formal experiment later on. The familiarization phase served as a preparation for subjects to get familiar with all the actions used in the subsequent test, and also experience the naturally occurring, but relatively small, body rotations in the movies (so as to minimize false alarms on the rotation task in the experiment).
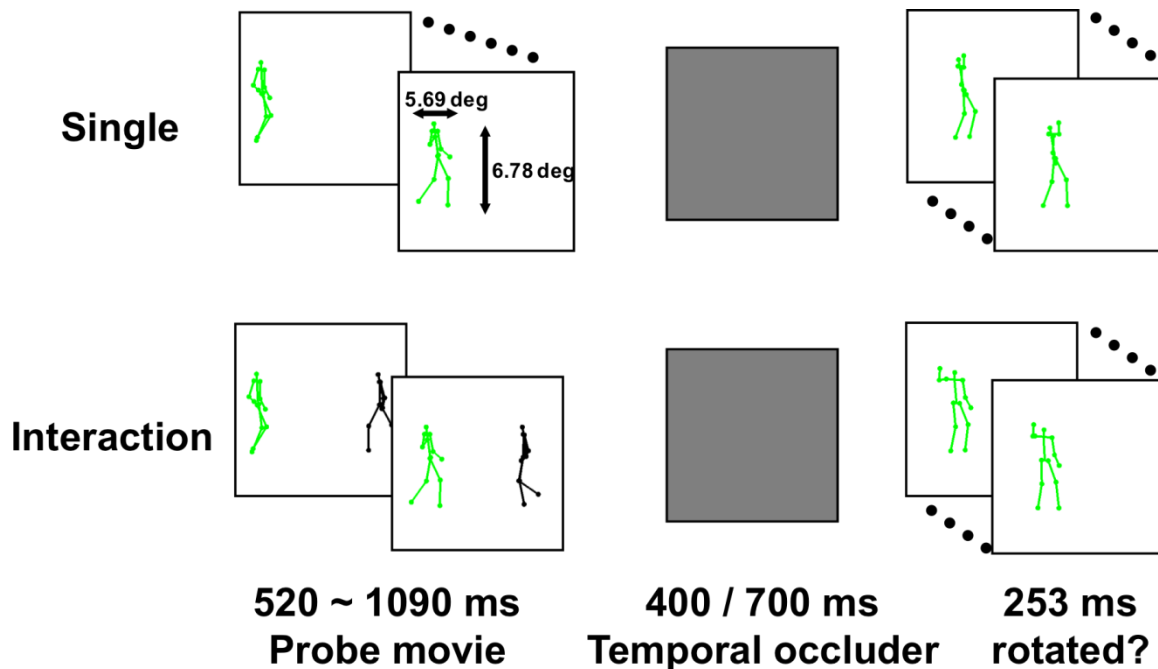
Figure 3-3. A sample trial for either single condition (upper) or interactive condition (lower) when the movie sequence is 'High-five'. The dynamic probe movie lasts for 520~1090ms for all the seven movie sequences (1667ms for 'High-five'). The temporal occlude lasts for 400ms or 700ms and the dynamic test stimulus lasts for 253ms.

3.2.2 Results

Subjects generally performed the task above chance levels (above 70% when the trials are easier, Figure 3-4). A repeated measure 3-way (Single vs. Interactive; Occluder 400ms vs. 700ms; Movement Gap 400ms vs. 700ms) ANOVA revealed a main effect of Movement Gap ($F(1,14)=13.323$, p=0.003), indicating that when the dynamic test figure was sampled from a time point close to the end frame of the probe movie, it is easier for subjects to tell the rotation of the action trajectory.

Next, we separate the data according to congruent or in congruent trials. In congruent conditions, a repeated measure ANOVA revealed no significant effects or interactions. However, in incongruent conditions, a repeated measure ANOVA revealed a significant main effect of Single vs. Interactive, indicating that the recognition accuracy was higher in the interaction condition than in the single-actor condition ($F(1,14)=5.342$, p=0.037). Thus, action prediction was more accurate when observing two interactive actors than observing a single actor, despite the temporal mismatch between action progression and time elapsed.

Subsequent paired t tests suggested a significant difference between the single and interactive conditions when the movement gap is 400ms in incongruent trials. The effect is 10.9% ($t(14)=2.534$, p=0.023). When the movement gap is 700ms in incongruent trials, the effect is not significant though, which is 12.1% ($t(14)=1.233$, p=0.238),
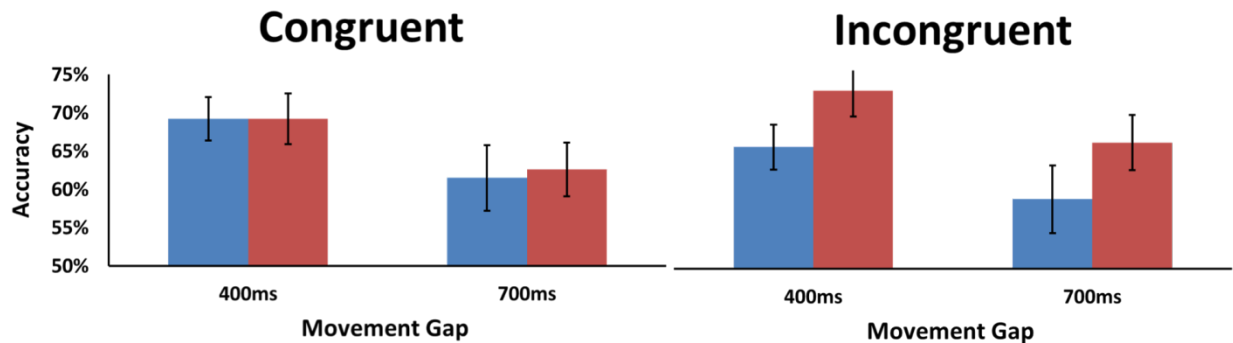
Figure 3-4. Accuracy on the rotation task. Blue bars represent single conditions. Red bars represent interactive conditions. The x-axis indicates 2 movement gaps (400ms and 700ms). The 4 bars on the left side are congruent trials (temporal occluders equal to movement gaps). The 4 bars on the right side are incongruent trials (temporal occluders don't equal to movement gaps). Stars above bars indicated significant paired t tests between single and interactive conditions within any movement gap condition in either congruent or incongruent trials.

3.2.3 Discussion

In Experiment 2, we hypothesized that the rotation detection task is presumable more sensitive to "what" information than "when" information, which should help interaction conditions over single-actor conditions. The results did support this view: in congruent conditions when the occluder time matches the movement gap (400ms or 700ms), no difference was found between the single-actor and the interaction conditions. However, more interestingly, in incongruent conditions when the occlude time mismatches the movement gap (i.e. the temporal precision is blur and irregular), recognition accuracy was significantly higher in the interaction condition than in the single-actor condition. Thus, action prediction was more accurate when observing interactions than observing a single actor, despite the temporal mismatch between action progression and time elapsed, showing that people obtained better performance in 'what' dimension of action prediction (i.e. predictive identification).

Experiment 2 did not obtain the results correspondent to Graf's study (2007) in which correctness in congruent trials are significantly higher than correctness in incongruent trials. This discrepancy may be due to the possibility that subjects are not sufficiently familiar with the new motion sequences we used here (we improved Experiment 1 by adding a session of familiarization which we designed and conducted after Experiment 2).

3.4 General Discussion

In this study, we investigated the difference of subjects' action prediction performance when there is a single actor compared to when there are two interactive actors with more social cues.

The first part of the results demonstrated that people can do real-time action prediction and give a good estimation of the action after 400ms temporal occluder. Actually, visual

perception is not simply a post-hoc reconstruction of the visual input of motion stimuli. In the 4th dimension other than the 3-D information, or the temporal dimension, plenty evidence demonstrated that visual motion perception is a predictive activity. For instance, flash lag effect involves the extrapolation of predictable movements (Devalois & Devalois, 1991). Representational momentum theory also showed that movement or implied movement of a physical object results in perceptual extrapolation of a trajectory beyond the actual object position presented on the screen, which distorted short-term memory later on (Bertamini, 1993; Finke, Freyd, & Shyl, 1985; Freyd & Finke, 1984). Neuroimaging studies support the perceptual extrapolation theory by showing that middle temporal (MT) area is activated even when the static visual stimuli imply future motion (Kourtzi & Kanwisher, 2000; Olson, Gatenby, Leung, Skudlarski, & Gore, 2004; Senior et al., 2000; Senior, Ward, & David, 2002). When these brain areas are deactivated by transcranial magnetic stimulation (TMS), the perceptual representational momentum effect disappeared (Senior et al., 2002). Thus, action perception is an online process by doing real-time simulation of the familiar daily human motions through utilizing the internal dynamic model.

It may be necessary to predict actions for motor control reasons. Several studies suggested a predictive function of the motor system in action prediction. To obtain fast and precise online motor control, the estimation of one's own body movement can be calculated ahead of the movement to solve the time delay problem, such as receptor transduction, neural conduction and central processing, especially in fast movement. Thus, to control one's own actions requires forward models which predict the sensory consequences of actions in real-time (Wolpert & Flanagan, 2001). It has been proposed that these already existing forward internal models can be also used to predict others' action as well (Blakemore & Frith, 2005; Grush, 2004; Prinz, 2006; Wolpert, Doya, & Kawato, 2003; Wolpert & Flanagan, 2001).

This dissociation between the 'what' and the 'when' dimensions is not limited to visual prediction. In human speech perception, there seem to be separate neurophysiological bases of sensory predictions of 'what' (predictive coding) which is defined as detecting predictive errors and using them to update internal template representations, and 'when' (predictive timing) which is defined as the internalization of temporal regularity or rhythms of events (Arnal & Giraud, in press). More specifically, in human speech perception, predicting 'when' (predictive timing) seems to be oscillation-based. That is to say, biological signals with certain low-frequency quasi-

periodic modulations can be aligned with slow endogenous cortical activity (Morillon et al., 2010). The resonance with neocortical delta-theta oscillations can be a plausible way to automate predictive timing as a low processing level (Andreou, Kashino, & Chait, 2011). However, when the brain tries to predict 'what' is going to happen according to the template matching, access to fine-grained spatial resolution is required which, in contrast, is not much necessary in predictive timing. Instead, 'what' prediction seems to be a hierarchical process when information flowing forwards through different stages of sensory regions is 'matched' by a top-down 'prediction' projected back from layer immediately above (Summerfield & Egner, 2009).

In the brain, visual and auditory temporal processing may share common neural mechanisms, probably the right parietal lobe (Battelli, Pascual-Leone, & Cavanagh, 2007). Speaking of temporal domain, there seems to be a parallel between apparent motion in visual modality and stream segregation in auditory modality. Moreover, the same brain area is involved in multisensory integration processing, such as the detection of synchrony between auditory and visual stimuli (Spence & Driver, 2000). Both neurophychological (Carlyon, Cusack, Foxton, & Robertson, 2001) and functional imaging studies (Cusack, 2005) suggested a metamodal computation imposed by right IPL(inferior parietal lobule) on any two consecutive evens, regardless of their sensory modality. Thus, temporal perception may be modality independent. The dissociation in human speech in auditory modality is probably produced by the same neural mechanisms underlying visual action prediction dissociation.

It could be more interesting if we can further investigate what is the underlying mechanism of processing interactive actions between two actors. Also, the separate neural substrates of multiple aspects of action prediction ('when' and 'what') need to be identified in brain imaging or electrophysiological studies to improve our understanding of the 4th dimension (time) of visual stimuli in this dynamic world.

Chapter 4 Flash lag effect of action joint prediction

The ability to localize moving joints of a person in action is crucial for interaction with other people in the environment. However, it is still unclear how the visual system encodes the position of joints in a moving body. We used a paradigm based on a well-known phenomenon, the *flash-lag* effect, to investigate the mechanisms underlying joint localization in bodily movements. When presented with a briefly-flashed joint during a walking action, observers perceived a strong flash-lag effect (i.e., when a briefly-flashed dot was presented physically in perfect alignment with a continuously moving limb, the moving joint was perceived as further ahead in the direction of the corresponding body movement). Our study revealed that the strength of this flash-lag effect of joint in biological motion depends on body orientation, suggesting that action-specific mechanisms are involved in determining the perceived position of moving joints. We also found that localization of joints was influenced by familiarity of actions, and by the functional importance of joints in a particular action. Simulation results based on the temporal averaging model (Krekelberg & Lappe, 2000) provided converging evidence that these empirical findings cannot be explained solely by generic motion mechanisms underlying the flash-lag effect. The present study provides compelling evidence that action processing interacts with position processing to localize the moving joints of whole-body actions.

The ability to localize a person and some body parts in action is crucial for action recognition and interaction with other people in the environment. For example, we avoid collision with an approaching pedestrian by estimating her moving body position in order to plan our own movements accordingly. In order to shake hands with a friend, we need to localize the position of her hand embedded within a moving arm. A couple performing ballet must accurately judge limb and body positions of their partner in order to perform well-coordinated dance movements. Although humans perform these tasks well, localizing the positions of joints nested within a moving body is not a trivial task. Due to intrinsic neural delay of position signal processing for moving objects, by the time the visual input arrives in cortical areas selective to biological motion, the joints and the body have already moved on to different positions. It still remains unclear how the visual system encodes the position of joints in a moving body.

In contrast to the dearth of research on joint localization in biological motion, there is a large body of research examining the mechanisms underlying localization of objects moving along simple movement trajectories (e.g., a moving bar in translation, a rotating line, a dot moving along a circle). When a briefly flashed object is presented physically in perfect alignment with a continuously moving object, observers perceive that the flash appears to lag behind the moving object (Mackay, 1958; Nijhawan, 1994). This well-known illusion, the *flash-lag* (FL) effect, provides a compelling demonstration that the visual system has developed mechanisms to cope with neuronal latencies in processing dynamic stimuli.

Two major classes of mechanisms have been proposed to account for the motion-induced position bias that constitutes the flash-lag illusion. The first class relies on generic spatial and temporal processing to either compensate for neuronal latencies, or reduce the delays in processing motion stimuli. For example, a model based on motion extrapolation assumes that the visual system makes spatial compensations by extrapolating the trajectory of a moving stimulus into the future, so that positions of the moving object are perceived to be ahead of the actual positions of the visual stimuli (Nijhawan, 1994). A differential latency model suggests that the visual system processes moving objects more quickly than briefly flashed stationary objects (Baldo & Klein, 1995; Purushothaman, Patel, Bedell, & Ogmen, 1998; Whitney, Murakami, & Cavanagh, 2000). This temporal difference yields the flash-lag percept, because the moving object has already shifted to a new position by the time the flashed object is processed.

A second class of mechanisms accounts for the flash-lag effect by assuming various forms of interaction between detection of the flash object and ongoing motion processing. For example, the temporal averaging model (Krekelberg & Lappe, 2000) suggests that the flash triggers temporal processing to integrate position signals of a moving object over a time window of 500 ms. Eagleman and Sejnowski (2000) proposed that the flash resets the window of motion integration processing, signaling the visual system to integrate post-flash position signals of the moving object. Although none of these mechanisms is sufficient to explain all the empirical findings related to the flash-lag effect, it is conceivable that multiple mechanisms could coexist and contribute differentially depending on stimulus complexity, object representation in space and time, and task demands (Whitney, 2002).

It is unclear, however, whether these mechanisms for extracting positions of moving objects with simple translation or rotation movements can account for the localization of

components in more complex motion patterns, such as bodily movements in human actions. Furthermore, if these generic mechanisms still apply to human body movements, how do these mechanisms interact with action representations to encode joint positions in a moving body?

Kessler, Gordon, Cessford, and Lages (2010) aimed to address these questions by examining the flash-lag effect with arm movements (e.g., moving an arm to reach out for a cup). These researchers confirmed the existence of the flash-lag effect for arm movements (i.e., the positions of a moving hand were perceived as lagging behind a flashed stationary object even when the two stimuli were spatially aligned). This finding generalized the FL effect to complex motion trajectories of a biological movement. Furthermore, the FL effect obtained when observing the original videos of arm movements was significantly greater than the effect measured in a control condition that showed symbolic moving shapes (i.e., removing human body appearance while maintaining the same motion trajectories). Hence, representing the motion stimulus as biological movement resulting from meaningful human actions enhances the FL effect. In addition, Kessler et al.'s study revealed that two other factors specific to human body movements (first-person perspective and sense of agency) modulate the magnitude of the FL effects when observing arm movements.

The present study aimed to extend previous research to investigate how the visual system encodes positions of moving joints when observing actions involving *whole-body* movements, and to determine how action representation interacts with generic mechanisms for localization of moving features to bias perceived positions of moving joints. To examine the mechanisms involved in encoding joint positions in a moving body, stationary joints were briefly flashed during the viewing of a walking action in a point-light display, and the magnitude of the FL effect was measured to assess the perceived location of moving joints. Given the dynamic nature of biological motion stimuli, we expected that participants would misperceive the relative position between flash joints and moving limbs in the action. To identify whether the FL effect depends on action processing, we used a critical control condition based on *inverted* actions. Perception of biological motion is known to be sensitive to body orientation; e.g., recognition performance is impaired if a point-light actor is presented upside-down (Sumi, 1984; Thurman & Lu, 2013). Since the motion profiles of joint trajectories are matched between upright and inverted actions, the role of action-specific mechanisms involved in localizing moving joints will be revealed by a difference in FL between the two conditions.

In addition, we examined how familiarity of an action, and the functional importance of joints in an action, impact the FL effect in biological motion. In order to specify the relative contribution of action processing in localizing joint positions, we compared human performance with a baseline model in which the FL effect was elicited solely by generic mechanisms in motion processing. To do so, we applied the temporal averaging model (Krekelberg & Lappe, 2000) to biological motion stimuli, allowing us to estimate the component of the FL effect that can be explained by general motion mechanisms, and to compare human performance with predictions derived from generic mechanisms based on the temporal averaging model.

4.1 Experiment 1: Joint flash-lag effect in biological motion

To investigate the mechanisms involved in localizing positions of moving joints, Experiment 1 used stimuli in which a stationary joint flashed briefly during the viewing of a walking action. We measured the flash-lag effect of two separate joints (hand and foot) in different body orientations of the actor (upright vs. inverted) performing a familiar or unfamiliar action (forward vs. backward walking). We mixed hand and foot joints in the experiment for two reasons. First, the randomized presentation order of flash joints prevented participants from tracking a particular joint movement when viewing the action. Tracking through smooth pursuit in eye movement can significantly influence the magnitude of the FL effect (Nijhawan, 2001). Using randomized flashes of different joints can minimize the strategy of tracking body parts, because particpants did not know where the dot would be flashed in a given trial. Second, the inclusion of two joints made it possible to examine whether localization of joints in a moving body depends on the functional importance of a joint to a particular action. Previous research has shown that the movements of foot joints play a more important role in discriminating bipedal actions (e.g., walking vs. running) than do hand joints (van Boxtel & Lu, 2012, 2015a). If joint localization varies with the functional importance of the joints, we would expect to observe a difference in FL between different joints.

4.1.1 Methods

4.1.1.1 Participants

The participants were 14 undergraduate students (9 females, M_{age}= 19.43 years) in the Psychology Department at the University of California, Los Angeles (UCLA). Participants reported normal or corrected-to-normal vision, and received two class credits for the 2-hour session required for the study. All studies in the present paper were approved by the UCLA IRB board. Consent forms and debriefing documentations were provided in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

4.1.1.2 Stimuli

The walker stimuli were retrieved from the CMU motion capture database (http:/mocap.cs.cmu.edu), and displayed from a sagittal view using the Biomotion Toolbox (van Boxtel & Lu, 2013a) in conjunction with the PsychToolbox (Brainard, 1997; Pelli, 1997). The moving walker was presented in the format of a skeleton display, consisted of 8 segments, including 4 limbs (i.e., two hands and two feet), 1 neck segment, 1 shoulder segment, 1 hip segment, and 1 torso segment formed by connecting the middle shoulder joint and the middle hip joint (see Figure 4-1). The skeleton walker was rendered as grey, except that one lower-arm and one lower-leg on the opposite side of the body were colored red, indicating that they were the limbs relevant to the task. The walker, located at the center of the screen, subtended about 4.39 degrees of visual angle in height and 2.01 degrees of visual angle in width. The walker completed one walking cycle (i.e., 60 frames) in 1 second. The extrinsic movements of the walker were removed to create the appearance of walking on a treadmill. The facing direction (either left or right) was randomly selected for a participant, and remained constant across the entire experiment for each participant. Facing directions were counterbalanced across participants. The walker was displayed with body orientation either upright or inverted and in walking directions either forward or backward, creating four conditions in the experiment. The backward-walking action was generated by playing the forward-walking video in reverse starting with an intermediate frame, so that the same posture was tested at the same time as in corresponding forward trials.

On each trial, a green dot in the size of 0.12 visual degrees was flashed for 33.3 ms at some point during the movement of the walker. The location of the flash dot was determined by the

reference joint (a foot or a hand joint) in frame 53, corresponding to 3/4 of the second step in the walking stimulus. The flash dot always appeared at the same physical location across all the conditions but at different time points. As shown in Figure 4-1, when the flash dot appeared several frames prior to, in the same frame as, or several frames after the time when the lower leg (or arm) reached the dot position, these stimuli were respectively labeled *behind*, *aligned*, and *ahead*, based upon the relative temporal relation between the position of the moving limb and the dot at the time of the flash. To minimize participants' confusion regarding which limb served as the reference for the judgment of relative spatial position, only the limb of interest (e.g., lower leg in Figure 4-1) was shown within the temporal window of +/- one frame around the time of presenting the flash.
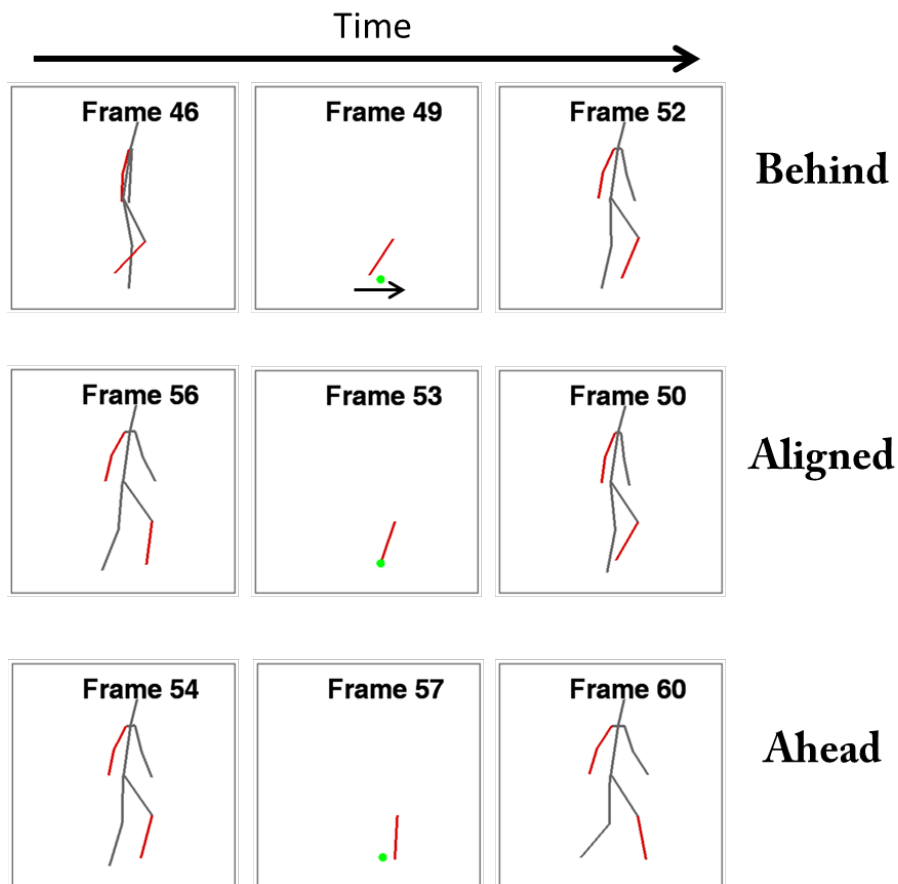
Figure 4-1. Stimulus illustration for Experiment 1. The moving limb (lower leg in this example) was physically behind, aligned with and ahead of the stationary flashed dot. The location of the flash dot was determined by the reference joint, either a foot or a hand joint. The flash dot always appeared at the same physical location across all the conditions but at

different time points. The arrow indicates the general moving direction of the reference limb around the time when the dot was flashed. See supplemental movies for the illustration, in which flash dot is perfectly aligned with moving body.

The temporal offsets between the flash dot and the moving walker varied in a range of -100 to +100 ms with a step size of 33.3 ms (i.e., 2 frames), resulting in seven levels of temporal offsets. Negative offsets indicate that the moving limb was temporally behind the flash dot, zero indicates perfect alignment between the flash dot and the moving limb, and positive offsets indicate that the moving limb was temporally ahead of the flash.

The experiments were conducted in a dark room. The stimuli were displayed on a calibrated Viewsonic CRT monitor with a refresh rate of 60 Hz and resolution of 1280 × 1024 pixels, and viewed from a constant distance of 45 cm maintained by a chin rest.

4.1.1.3 Procedure

The experimental design involved three within-subject factors: reference joint (hand vs. foot), walking direction (forward vs. backward) and body orientation (upright vs. inverted). Observers participated in one session on the first day and a second session after a week, with either upright or inverted walkers being shown in each session. A session consisted of one block with forward walking direction and another block with backward walking direction. The order of walking direction and body orientation were counterbalanced across participants. In each block of 280 trials (2 joints × 7 offset levels × 20 trials per condition), hand/foot reference joints were randomly intermixed.

On each trial, participants were asked to judge whether the flash dot appeared on the left or the right side of the red limb. Due to the brief presentation of the flash dot, participants were also allowed to report that they did not observe the flash. Responses were re-coded into whether the moving limb was perceived ahead of or behind the flash dot according to the motion trajectory of each reference joint around the location of the flash. Participants completed 10 practice trials prior to the experiment without feedback. The presentation of the practice trials was twice as slow as the experimental trials in order to show participants the brief presentation of the flash dot.

4.1.2 Results

The analysis was based on the mean proportion of trials on which the moving reference limb was reported ahead of the flash dot as a function of spatial offsets between the flash and reference joint location of the walker at the time of flash. To quantify the strength of FL in each condition, we measured the shift of the point of subjective equality (PSE), which reveals the average spatial displacement at which participants would yield 50% "ahead" responses when indicating the perceived alignment between the moving limb and the flash dot. Trials on which participants reported failure to observe a flash dot were removed from data analyses. Overall, the miss rate was very low (mean proportion of 1.3%).

For individual observers in each condition, we fitted the data with a Gaussian cumulative distribution function with two parameters, mean μ and standard deviation σ, as shown in Figure 4-2 (top panel). For all observers, this function fitted their data well (R^2 of .946 ± .075). The estimate of the μ parameter was taken as the PSE estimate, corresponding to the spatial offset yielding 50% "ahead" responses, and provided the measure of the strength of the FL (perceived spatial lag). A PSE of zero indicates absence of a flash-lag effect (i.e., participants yielded 50% "ahead" responses when the flash dot aligned with the moving reference limb), and a positive PSE indicates presence of an FL effect (i.e., a moving limb that spatially lagged behind was perceived in alignment with the flash).
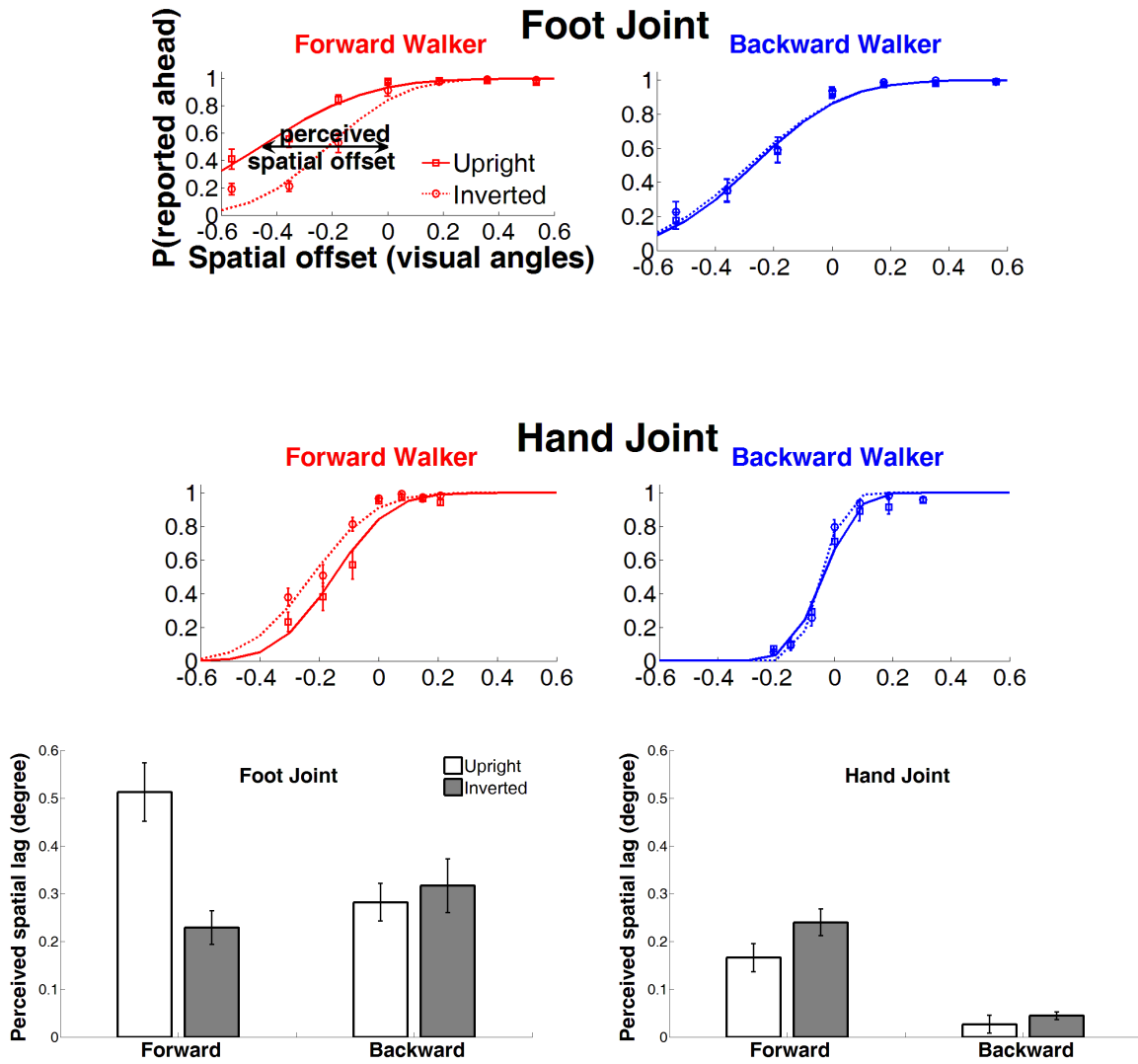
Figure 4-2. Results for joint FL effect in Experiment 1. Top panel: the fitted psychometric functions based on the average performance for different joints and walking directions. Bottom panel: The strength of FL in terms of the perceived temporal lag of PSE varied depending on body orientation (upright vs. inverted), walking direction (forward vs backward) and reference joints (foot vs. hand). The error bars indicate SEM in all the plots of the paper.

As shown in Figure 4-2 (bottom panel), flash-lag effects (i.e., positive PSE values) were observed in most conditions. A repeated measures ANOVA with three within-subject factors (foot vs. hand joint; body orientation, and walking direction) revealed a significant three-way interaction effect ($F(1,13) = 18.97$, $p = .001$, $\eta_p^2 = .59$), suggesting the impact of body orientation on localizing joint position in a walking body depends on specific joints and familiarity to the

walking action. Specifically, when examining the localization of foot joint by flashing the dot close to the lower leg, we found a significant two-way interaction ($F(1,13)=20.04$, $p = .001$, $\eta_p^2 = .61$) between body orientation and walking direction. In contrast, for perceived location of moving hand joints, there was no significant interaction ($p = .208$) between body orientation and walking direction.

The perceived location lag of the foot joint in the upright forward walker was significantly stronger (mean = 0.51 degrees) than in the inverted forward walker (mean = 0.23 degrees) ($t(1,13)=4.75$, $p < .001$). But such a difference for the foot joint induced by body orientation was not observed for the unfamiliar action of backward walking ($p = .526$). These results suggest that when observing an unfamiliar action such as backward walking, FL may be mainly elicited by generic mechanisms based on motion trajectories of individual joints, which is not sensitive to body orientation in the whole-body action. However, for a familiar action (i.e., forward walking), FL appears to depend on two mechanisms: action-specific mechanisms that are sensitive to body orientation, and general motion mechanisms that depend on specific motion trajectory of joints.

When the flash dot was located at the hand joint, significant flash-lag effects were observed in most conditions ($ps < .01$ except in the upright backward walker condition), indicating a perceived spatial lag of the hand joint in the moving body. A much stronger FL effect was found in the forward-walking than in the backward-walking action ($F(1,13)=39.63$, $p < .001$, $\eta_p^2 = .75$), consistent with the finding with foot joints that action familiarity impacts the perceived mislocalization of moving joints. We also found a main effect of body orientation ($F(1,13)=10.69$, $p = .006$, $\eta_p^2 = .45$), indicating stronger FL of hand joints in the inverted walker than in the upright walker. This difference was in contrast to the results found for the foot joint. For a forward walking action, the flash lag effect of hand joint for the inverted walker (mean = 0.24 degrees) was larger than for the upright walker (mean = 0.17 degrees), $t(13) = 2.43$, $p = .030$. We speculate that multiple factors may contribute to this difference. First, a spatial attention preference may be applied to the closer-to-ground body parts in bipedal action (such as walking, running), which constitute informative cues for a life-detector mechanism based on foot movements (Troje & Westhoff, 2006). When the body orientation is upside-down, the hand joints are placed in relatively lower positions than foot joints. Second, as shown in Figure 4-3, the hand motions of the inverted walker reveal vertical velocity and acceleration profiles consistent with the foot motions of an upright walker (e.g., lifting upward then downward),

70

which are important cues for signaling biological movements (Chang & Troje, 2009a). Hence, for an inverted walking action, the hand motion in inverted (as compared to upright) walking may provide more informative signals specialized to biological movements, resulting in a stronger FL effect for the hand joint in the inverted condition than in the upright condition.
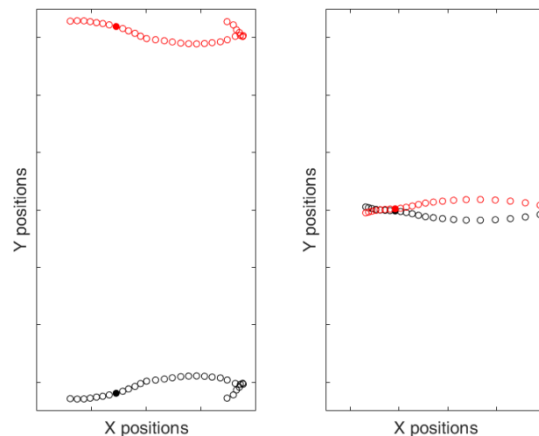
Figure 4-3. Illustration of motion trajectories. The two panels depict the movements of foot and hand in the upright (black) and inverted (red) conditions, respectively. The circles indicate positions of each joint when moving from right to left. The solid circles indicate the position of the flash dot in Experiment 1. When the body orientation is inverted, the hand motion shows similar vertical motion profiles (moving upward then downward) as the foot motion in the upright walking condition.

In summary, Experiment 1 demonstrated the presence of a flash-lag effect in biological motion for localizing the moving joints of whole-body actions. An alternative way of comparing the flash-lag effects between conditions is to measure the lag effect in units of time, rather than spatial lag as shown in Figure 4-2. We converted the FL effect into time units, and found the mean temporal lags for the foot joint were 94 ms for upright-forward condition, 43 ms for inverted-forward, 52 ms for upright-backward foot, and 59 ms for inverted-backward foot. For the hand joint, we found lags of 57 ms, 82 ms, 14 ms and 19 ms for the four corresponding conditions. Overall, the two units of measurement (time and space) yielded similar results.

We found that the magnitude of FL effects depended on body orientation of the observed actions (upright vs. inverted walkers), familiarity of the actions (forward vs. backward walkers), and joints (foot vs. hand joints). Among these three factors that were manipulated, motion trajectories were well-matched between the upright and inverted conditions, as inversion of body orientation maintains other motion signals (e.g., speed, velocity). Hence, the difference between

upright and inverted forward walking provides strong evidence supporting the impact of action representation on the localization of moving joints in the body.

However, for the other two factors (familiarity and joints), each condition has its own specific dynamics in motion trajectories. For example, hand and foot joints moved at different speeds, and the movements before and after the flash differed between the forward and backward walking conditions. To quantify how low-level motion signals influence the strength of the FL effect, we employed the well-established temporal averaging model (Krekelberg & Lappe, 2000) to estimate the misperceived location of moving joints based solely on their individual motion trajectory around the time of the flash. The model simulation results are presented in the next section.

4.1.3 Model simulation results using the temporal averaging model

The temporal averaging model (Krekelberg & Lappe, 2000) processes motion input with a slow temporal filter and two spatial filters to compute local motion energy, and predicts the perceived position of a moving target to be the location corresponding to the maximum motion energy. To account for the FL effect, the model calculates the relative distance between the moving target and a flash as the temporal average of the position difference between the moving object and the stationary location of the flash over a time window of 500 ms.

We applied the temporal avearging model to quantify the flash-lag effect elicited by general motion mechanisms (as opposed to action-specific mechanisms). To adopt the model for the walking stimuli used in the present study, two modifications were made. First, the spatial filters only analyzed horizontal movements from an input stimuli consisting of either hand or foot movements. Participants' task was to judge whether the moving limb was on the right or left side of the flash dot; hence only perceived horizontal positions were relevant to the task. Accordingly, an analysis based on 1D horizontal movements would produce the same decision as a model based on 2D spatial-temporal filters. Second, the dot was flashed around 3/4 of the duration of the second step in a walking cycle lasting 1 second. The interval from the flash onset to the end of the motion stimuli was less than 500 ms (the temporal window size used in the original temporal averaging model). Our simulations used a temporal averaging window determined by the time interval between the flash onset and the end of the walking sequence. In our setup, the maximum integration duration was 238 ms (in the condition with temporal offset of -100 ms).

The simulation results from the temporal averaging model are depicted in Figure 4-4. The model provides estimates of the perceived spatial displacement between the moving joint and the dot flashed at any time point during the walking stimulus. The FL effect was measured as the displacement at which the model decided the flash in alignment with reference limb. Since the perceived spatial displacement followed a linear relation as the function of the physical spatial distance between the moving joint and the flash (linear fit $R^2 > .99$), a linear regression was used to estimate the FL effect. As shown in Figure 4-4, the model predicts that the FL effect will not vary in magnitude when body orientation is changed from upright to inverted walking. This prediction implies that the general mechanism underlying the FL effect solely relies on joint trajectories, and does not interact with action representations.
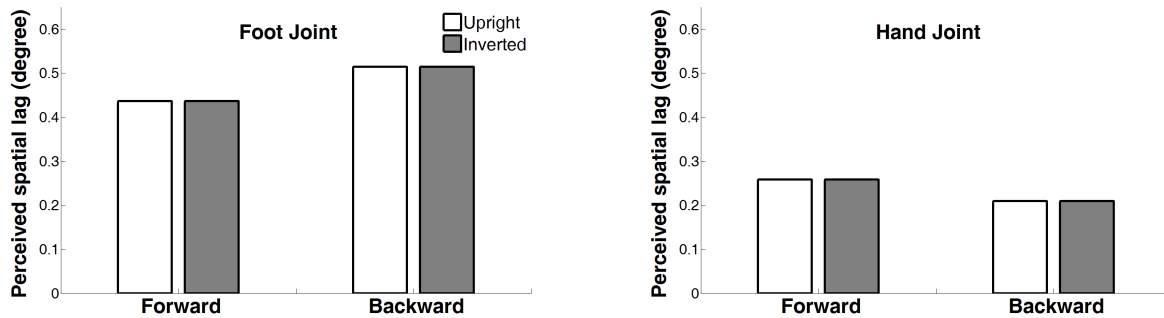
Figure 4-4. Temporal averaging model predicts the flash-lag effect for foot and hand joints. The flash-lag effect is predicted to remain the same when the walking sequence is inverted, but to vary as a function of walking direction (forward vs. backward).

However, there were differences in model predictions for forward versus backward walking. For example, for the foot joint, the model predicts a larger FL effect for backward than forward walking. In contrast, the model predicts that hand joints will elicit a greater effect in forward than backward walking. We expect that these predicted differences are due to the change of motion trajectories around the flash time in the stimuli for the two walking directions. To test these predictions derived from the temporal averaging model using generic motion processes, we conducted Experiment 2 to control the visual input so that we could remove the impact of action representations on the FL effect.

4.2 Experiment 2: Flash-lag effect in the absence of action perception

In Experiment 1, we demonstrated the existence of a FL effect for joint movements in biological motion, and also established the impact of body orientation on the misperception of joint positions. If these findings indeed signal the involvement of action processing in the flash-lag effect, then the impact of body orientation on FL should disappear if the action is not perceived (e.g., if the movement of reference limb is shown without the other body parts in the display). Experiment 2 aimed to test this hypothesis by presenting the same reference limbs and flash dot as in the previous experiment, but in the absence of other body parts. Hence, we expect that the pattern of FL effects obtained in Experiment 2 will be well predicted by the temporal averaging model.

4.2.1 Methods

4.2.1.1 Participant

Twelve undergraduate students (8 females, 4 males, M_{age}= 20.92 years) in the UCLA Psychology Department (who had not participated in Experiment 1) participated in this experiment to receive two class credits. All participants reported normal or corrected-to-normal vision. Data for two participants were removed from the analysis due to poor psychometric curve fitting (i.e., R^2 less than .7).

4.2.1.2 Stimuli and procedure

The stimuli in Experiment 2 were the same as those in Experiment 1 except that only the two reference limbs were shown in the display without presenting other parts of the body. Specifically, only one lower leg and one lower arm were shown on the screen. Because the other body parts were absent, participants typically perceived the stimuli as two bars moving as pendulums, and were not aware that these movements were derived from human walking actions. Figure 4-5 illustrates the stimuli used in the limb-only display. The procedure was identical to that of Experiment 1.
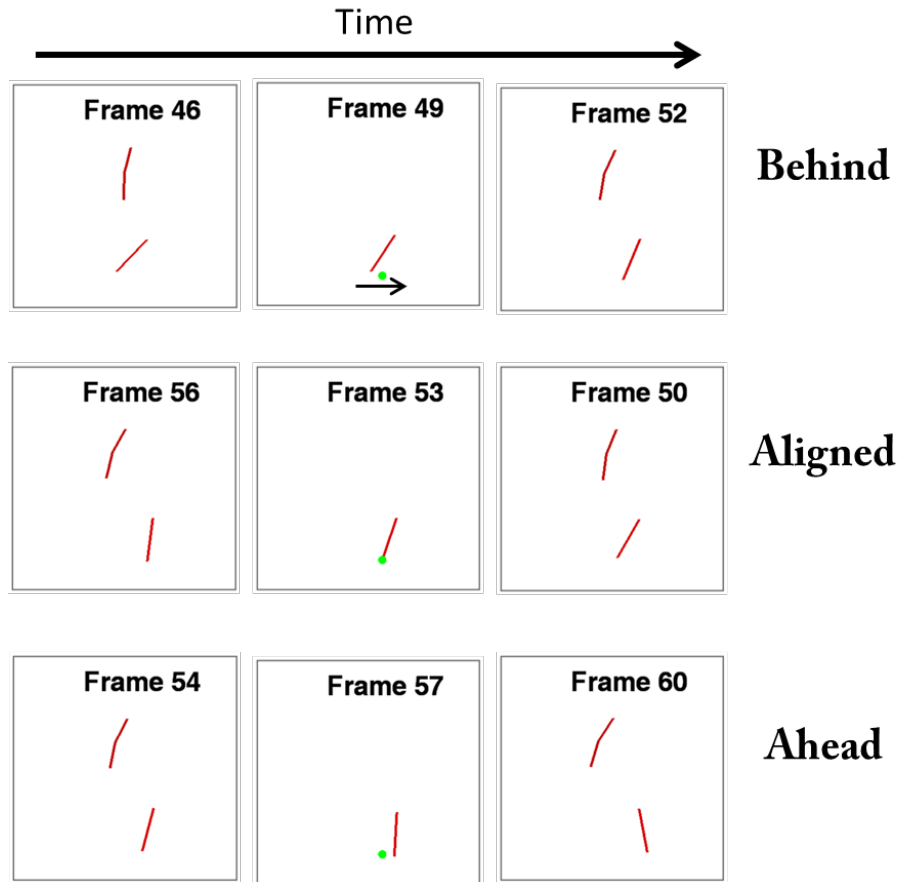
Figure 4-5. Illustration of the display used in Experiment 2. The moving reference line was physically behind, aligned with or ahead of the stationary flashed dot. The arrow indicates the general moving direction of the line around the time when flashing the dot.

4.2.2 Results

Similar to the results of Experiment 1, a FL effect was still observed for the limb-only display, as evidenced by the significant spatial lag perceived by participants in most conditions ($ps < .01$ except in upright backward walker condition) (see Figure 4-6). However, there was no difference between the upright and inverted conditions for the limb-only display. Since Experiment 2 did not provide sufficient input information to recruit action processes, localization of moving joint were not influenced by the body orientation, as was predicted. A repeated-measures ANOVA revealed a significant main effect of joint ($F(1,9) = 60.45, p < .001$), showing a greater FL effect for the foot joint than the hand joint. This difference can be explained by the difference of movement speeds between the two joints: foot joints moved such faster than hand joint in the walking action, as predicted by the temporal averaging model (see model predictions

in Figure 4-4). We did not find significant a three-way interaction between the three factors (i.e., joint, walking direction and body orientation), in contrast to the result observed in Experiment 1. The limb-only display in Experiment 2 also showed a significant two-way interaction of joint and walking direction, revealing that the FL strength for foot joint in backward walking was greater than the effect in forward walking, whereas the hand joint showed an opposite pattern ($F(1,9)$=14.75, $p = .004$, $\eta_p^2 = .62$). Note that this interaction effect is predicted by the temporal averaging model (see Figure 4-4). Overall, the results of Experiment 2 were qualitatively consistent with the generic mechanism model based on temporal averaging. These results further confirmed that when whole-body movements were absent in the display, participants fell back on the default of using generic motion mechanisms to perceive the location of a moving segment.

Comparing the findings of the first two experiments, for the forward walking action the strength of FL for foot joint in the limb-only display of Experiment 2 (0.17 degrees) was significantly reduced relative to to the strength of FL when the whole-body action was observed in Experiment 1 (0.28 degrees, $t(22)$= 4.49, $p < .01$). When we analyzed the results from Experiments 1 and 2 in a four-way mixed ANOVA using the appearance of body configuration (presence or absence of the body) as a between-subject factor, together with the other three within-subject factors, we found a significant four-way interaction ($F(1,22)$=14.38, $p = .001$, $\eta_p^2 = .40$). This result was driven by a significant three-way interaction of body orientation, walking direction and joints in Experiment 1 with the whole-body display, but lack of such an interaction effect in Experiment 2 with the limb-only display. Hence, the limb-only display effectively removed the modulation of body orientation (upright vs. inverted) on the FL.
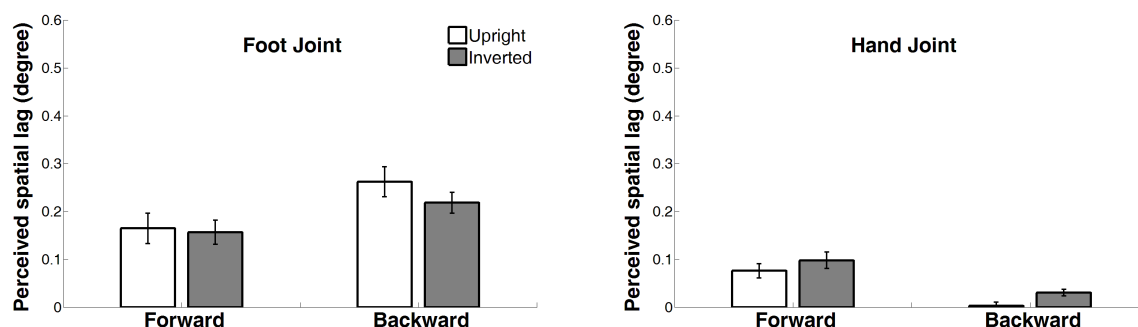
Figure 4-6. Results of joint FL effect in Experiment 2. When only relevant limbs were presented, the strength of FL was significantly weakened, and was not influenced by whether the reference limbs were from an upright or inverted walker.

Since the procedure of Experiment 2 did not elicit involvement of action processing in the localization of joints, the strength of the flash-lag effect measured here is informative for estimating the time window for integrating perceived positions after the flash. We considered the integration window as a free parameter in the temporal averaging model by Krekelberg and Lappe (2000), and searched for the best-fitting parameter to account for human performance across all the four upright conditions (foot forward, foot backward, hand forward, hand backward) in Experiment 2. We found that a temporal integration window of 115 ms after the flash yielded the best model predictions (.16, .21, .10, .07) for the four conditions to fit the average performance of humans, (.16, .24, .08, .02, respectively). The root mean square deviation (RMSD) between model prediction and human performance was .001, and the correlation was .99.

Note that the best-fitted estimate of the temporal window was much shorter than the temporal integration window of 500 ms proposed in the study of Krekelberg and Lappe (2000). This difference can be explained by multiple factors. First, movements of joints in human actions involve accelerations and oscillated movements, which may evoke a different temporal integration window than does motion with constant velocity, as used in most previous studies. Second, in our stimulus the time of showing the flash was close to the end of the motion sequence, followed by a rapid disappearance of the stimulus, which may have led to a shortened integration window given the limited input of dynamic information. The best-fitting temporal window (115 ms) in our study is in agreement with the temporal window derived in some previous studies. For example, a study using rotation stimuli derived a temporal range of 80 to 100 ms underlying the observed flash-lag effect (Eagleman & Sejnowski, 2000). Furthermore, in a recent study using transcranial magnetic stimulation (TMS), the researchers used translation stimuli to show that the maximum effect for TMS pulses in reducing the flash-lag effect occurs 200 ms after the flash (Maus, Ward, Nijhawan, & Whitney, 2013). This estimate is consistent with the value obtained in our simulation results.

4.3 Experiment 3. Flash-lead effect in biological motion stops at the instant the flash occurs

In the previous two experiments, we found robust flash-lag effects in biological motion, such that a moving joint aligned with a flash was seen as further ahead in the direction of the corresponding limb movement. Although flash-lag effects are commonly observed in many different situations involving moving objects, the effect can be abolished or even reversed in some conditions. For example, if the moving object stops motion at the time of the flash, the flash-lag effect is not observed; instead the flash and moving object are perceived to be aligned (Eagleman & Sejnowski, 2000; Kanai, Sheth, & Shimojo, 2004; Nijhawan, 2001). In some cases, an opposite effect, flash-*lead*, is observed such that perceived position of the moving object at its termination undershoots its physical position (Krekelberg & Lappe, 2000; Maus & Nijhawan, 2009; Patel, Ogmen, Bedell, & Sampath, 2000; Roulston, Self, & Zeki, 2006). Experiment 3 was designed to test the flash-termination conditions to examine whether action representation still interacts with localization processing of moving joints when human actions terminate at the instant the flash occurs.

4.3.1 Methods

4.3.1.1 Participants

Thirteen undergraduate students (7 females, M_{age}= 20.92 years) in the Psychology Department at UCLA participated in this experiment. Participants reported normal or corrected-to-normal vision, and received one class credit for the 1-hour session required for the study.

4.3.1.2 Stimuli

The stimuli were the same as the forward walking session Experiment 1, except that the walking stimuli disappeared right after the dot flashed on the screen, as shown in Figure 4-7. Experiment 3 did not include the backward walking conditions. Note that the stimulus manipulation in Experiment 3 differed from some flash-termination stimuli in previous studies (e.g. Eagleman & Sejnowski, 2000) in that the moving limbs disappeared at the end of its movement, rather than stopping. Participants were asked to judge whether the flash dot was located on the left or right side of the reference limb.
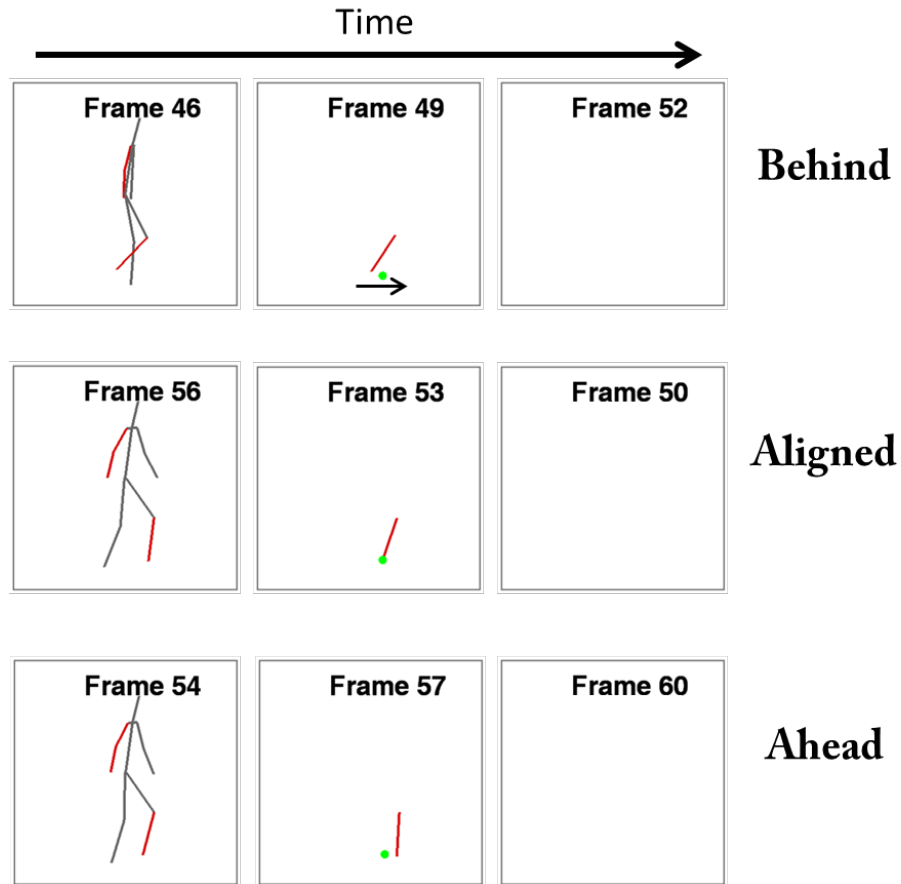
Figure 4-7. Illustration of the display used in Experiment 3. The flash dot appeared at the same spatial location but at different time points across all conditions. When the dot flashed, the moving reference limb was either physically behind, aligned with or ahead of the stationary flashed dot in three conditions. The arrow indicates the general moving direction of the line around the time when flashing the dot. The walker disappeared right after the dot flashed.

4.3.2 Results

Figure 4-8 shows the negative values for perceived spatial lag. The negative values indicate a flash-lead effect in which the moving joint is perceived to lag behind the stationary flashed dot even when they were physically aligned. This result is consistent with previous findings that participants misperceive the final position of a moving limb prior to its disappearance in the direction opposite to its motion trajectory (Maus & Nijhawan, 2009; Roulston et al., 2006). A repeated-measures ANOVA revealed only one significant effect, the main effect of joint ($F(1,12)=5.95$, $p = .031$, $\eta_p^2 = .33$),indicating a larger flesh-lead effect for the foot than the hand joint. This difference is likely due to the different movement speeds of the two

joints, consistent with the previous finding of a speed dependency for the strength of the flash-lead effect (Roulston et al., 2006).

Importantly, Experiment 3 did not reveal any difference in the flash-lead effect between upright and inverted actions. This result contrasts with the finding of a body orientation dependency for the flash-lag effect in Experiment 1. A mixed ANOVA confirmed this difference, yielding a significant three-way interaction when the two experiments were analyzed together ($F(1,25)=15.81$, $p = .001$, $\eta_p^2 = .39$). The disappearance of the impact of body orientation on localization of moving joints in the flash-termination condition suggests that the interaction between action representation and location perception of moving joints likely takes place at or after the flash event.

The flash-lead effect revealed in Experiment 3 can be explained by slightly modifying the temporal averaging model (Krekelberg & Lappe, 2000). The localization of a moving joint can be derived by averaging its perceived position signals over a time window either prior to or at the flash event. Perceived location of a movement joint in the flash-termination condition can be calculated by disabling the last stage in the temporal averaging model. When the action stimuli suddenly disappear, a strong retinal transient signal could be elicited to possibly suppress the integration stage. In other words, the perceived positions could be directly derived at the time point when the flash occurred with the predicted spatial lag PSE of -.16, -.11, -.03, -.06, respectively, for the foot-forward, foot-backward, hand-forward and hand-backward conditions. These predictions are qualitatively consistent with human performance.
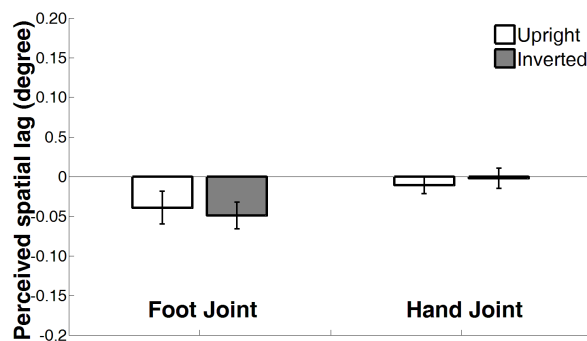
Figure 4-8. Results of Experiment 3: perceived spatial offsets in the flash-termination display with forward walking. The negative values of perceived spatial lag indicate that the moving joint was perceived to lag behind the stationary flashed dot even when they were physically aligned.

4.4 General Discussion

The results of the present study provided evidence that people misjudge the locations of joints in a moving body. Action representations interacts with generic motion mechanisms underlying position processing of dynamic stimuli to evoke the misperception of joint location in a moving body. First, we found robust flash-lag effects for localizing joints in a walking action, supporting the hypothesis that mislocalizing spatial positions of moving objects in dynamic stimuli results from a primitive operation of the visual system. Based on simulation results from the temporal averaging model obtained by sampling positions in a temporal window, we showed that the model can explain the presence and absence of a flash-lag effect, as well as changes in the size of the effect across different joints (basically due the faster movements in foot joints than hand joints).

However, this generic motion mechanism of sampling positions in a temporal window cannot account for the key finding in the present study: the magnitude of the FL was influenced substantially by the characteristics of the action. As shown in Experiment 1, observing a walker with a natural body orientation (i.e., upright) yielded a significantly stronger FL effect for the critical foot joint than did viewing an inverted walker. This difference cannot be attributed to general motion extrapolation based on low-level motion cues (e.g., velocity, direction, or acceleration) or temporal averaging mechanisms of perceived position signals, since the characteristics of foot movement trajectories in the upright and inverted conditions were well matched. The difference in FL effects for upright and inverted walkers disappeared when the actor performed backward walking, an unfamiliar action involving body movements that are unusual in everyday life. Hence, action characteristics such as familiarity modulate the strength of FL to influence the perceived location of joints in body movement.

Experiment 2 showed that the impact of body orientation on the FL results from an interaction between action representation and the processing required to determine the instantaneous position of a continuously moving object. When body parts were absent and only a

moving line segment of a reference limb was shown in the display, the flash-lag effect was still observed, but the impact of body orientation on the FL disappeared.

Experiment 3 showed a flash-lead effect, a different type of mislocalization of moving joints, in which the perceived position of the moving joint at its termination undershoots its physical position. However, the flash-lead effect obtained in Experiment 3 with the flash-termination display did not show an impact of body orientation on the misperceived location of moving joints in biological motion. In comparison with the main finding in Experiment 1, the lack of an influence of body orientation on the flash-lead effect in Experiment 3 implies that the interaction between action representation and generic mechanisms for determining the perceived position of a moving joint likely takes place *after* the occurrence of the flash event.

There are apparent functional benefits of perceiving the location of joints in a moving body biased toward the direction of future motion, as revealed by the flash-lag effect in actions reported in the present paper. When a friend walks towards you to shake hands, it is better to overshoot you position estimation so that you put your hand in a future possible location to intersect with his hand movements, rather than underestimating the moving hand position so that you miss his hand altogether. Moreover, our finding that action processing interacts with motion mechanisms for localizing a moving joint are consistent with previous research using relatively simple movement patterns, which suggests an interaction between localization of a moving object and the motion configuration of the object (Kohler, Cavanagh, & Tse, 2015; Whitney & Cavanagh, 2000). For example, Whitney and Cavanagh (2000) showed that the configuration of motion in the visual input impacts the perceived locations of both moving and stationary stimuli, even when the motion patterns are at a substantial distance from the stimuli.

The finding that action representations modulate the localization of moving joints also agrees with findings from a study by Kessler et al. (2010), which showed that first-person perspective and sense of agency can modulate the magnitude of the FL to bias the perceived hand positions in arm movements. Interactions between higher-level processing (i.e., action) and lower-level processing (i.e., localization of a moving element) have also been found between the visual and the sensorimotor system, as active control of the moving stimulus by the participants can impact the magnitude of the flash-lag effect (Ichikawa & Masakura, 2006, 2010; Nijhawan & Kirschfeld, 2003; Scocchia, Grosso, de'Sperati, Stucchi, & Baud-Bovy, 2009).

The simulation results reported in the present paper are based on the well-known temporal averaging model developed by Lappe and Krekelberg (1998). This model assumes that the perceived position of a moving target can be derived from motion processing (i.e., the location corresponding to the maximum motion energy). In order to read out the instantaneous position of a moving object at a time point, the perceived position signals are integrated within a temporal window. Such computation to extract position information from motion seems reasonble when we deal with more complex motion stimuli. For human actions involving a dozen moving elements, it would be costly to encode the positions of all the joints at each time frame, since most tasks in action recognition may not invovle judgments regarding the specific joint positions. However, when a task requires to localizing moving joint(s), sampling perceived position information within a temporal window can be performed.

How does action representation interact with the generic mechanisms underlying localization of moving objects? There are several possible ways to extend the computational subcomponents involved in sampling position information within a temporal window. As shown in Figure 4-9 (left), one possibility is that the perceived positions that feed into the temporal integration stage are based on *predicted* body movements derived by action simulation processing. Previous research using a priming and representation-momentum paradigm revealed that observers automatically anticipate the future posture of observed actors (Jarraya et al., 2005; Kourtzi & Shiffrar, 1999; Thornton & Hayes, 2004; Verfaillie & Daems, 2002). Recent studies have shown that the visual system is able to predict future postures almost in real-time, supporting the hypothesis that a simulation process operates to enable the prediction of future actions (Graf et al., 2007; Manera, Schouten, Verfaillie, & Becchio, 2013; Sparenberg et al., 2012; Springer & Prinz, 2010).

If action prediction takes place for familiar and natural body movements (such as foot movements in an upright action), we would expect that this extension to the future could effectively elicit more spatial extrapolation of the perceived positions in the upright condition (as shown in the red dashed line in Figure 4-9, left plot), than in the inverted condition (as shown in the red dotted line in Figure 4-9, left plot). Action processing employs prediction-by-simulation mechanisms to extrapolate body movements towards the future in the upright condition, yielding a strong FL effect in Experiment 1. Weaker prediction results in less spatial extrapolation in the

inverted condition, leading to a reduced FL effect. In Experiment 2, the lack of action prediction in the limb-only condition may have led to the absence of a FL effect.

One could argue that the results of our Experiment 3, showing a flash-lead effect in the flash-termination condition, provide evidence against this mechanism. However, the sudden disappearance of action stimuli may send a strong signal to terminate the process of motion extrapolation, as suggested by the results of several previous studies (Maus & Nijhawan, 2008, 2009; Nijhawan, 2008). The visual system may stop the process of prediction-by-simulation because the probability of interacting with this person is very low. Though some lingering extrapolated actions after the termination signal may need time to take place, the integrated position signals within the temporal averaging window would likely be more biased towards the past.

An alternative mechanism by which action processing could interact with localization processing is to vary the temporal integration window in different conditions, as shown in Figure 4-9, right plot. If the observed body movements are natural (e.g., with upright rather than inverted body orientation) and familiar, a larger temporal window may be used for temporally averaging perceived location signals, relative to unfamiliar and unnatural actions. As shown in Figure 4-9, right plot, the dependency of foot-joint FL effect on body orientation (upright vs. inverted) in Experiment 1 can be explained by a longer temporal integration window imposed for the upright forward walker than for the inverted walker conditions. Why would a larger temporal window be employed for familiar actions with upright body orientation? Wang and Jiang (2012) found that an upright walking sequence was subjectively perceived as longer than an inverted walker, and the same temporal dilation effect was found even when participants viewed movements of only one joint (e.g., foot) without awareness of the walking action. The lengthened temporal duration in the upright walking action may increase the integration window size in subjective time. as proposed in a computational account offered by Rao, Eagleman, and Sejnowski (2001).
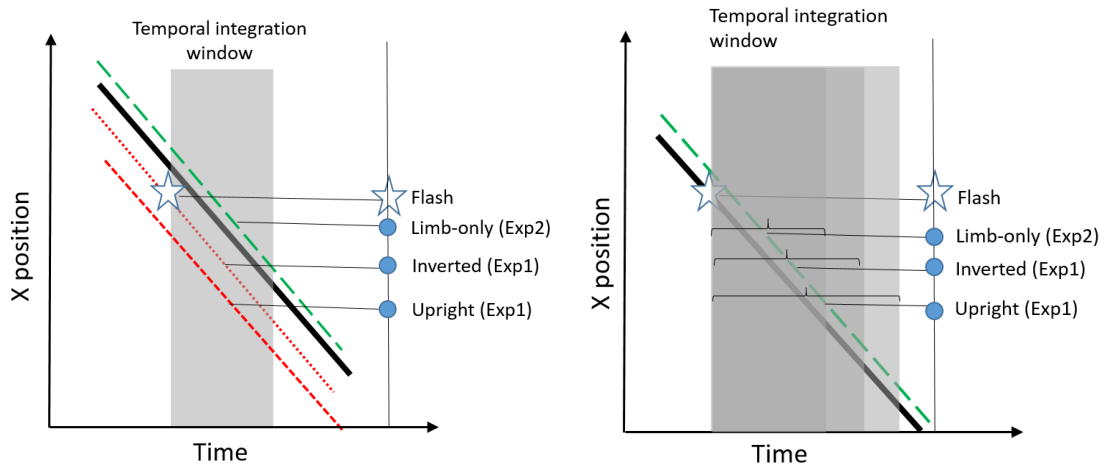
Figure 4-9. Illustrations of two possible mechanisms underlying the interaction between action processing and localization of a moving joint. Left, action processing employs prediction-by-simulation mechanisms to extrapolate body movements towards the future in the upright condition to yield a strong FL effect in Experiment 1. Weaker prediction results in less spatial extrapolation in the inverted condition, reducing the FL effect. The lack of action prediction in the limb-only condition results in the absence of an FL effect in Experiment 2. Right, action processing may vary the subjective time duration for biological movements, which turn into a longer integration window for natural and familiar actions (such as the upright condition) than for other conditions in the experiment.

The flash-lag effect affords a window to study the interactions between high-level representation of human movements and low-level motion and position processing. More studies are needed to further pin down the interaction between action processing and different mechanisms for perceiving the locations of moving joints and the body. Predicting the location of bodily movements may occur at different time scales, from a fraction of a second for predicting simple movements, to minutes or even hours for understanding dynamic and social events. In the present paper, we examined the mechanisms underlying perception of the location of a moving body in the immediate future (i.e., in the order of a fraction of a second). This type of real-time prediction may be closely connected with the unconscious and automatic tendency to imitate others' behavior (Chartrand & Bargh, 1999; Wilson & Knoblich, 2005), likely by recruiting the mirror neuron system to predict immediate future actions (Dipellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992). To predict actions over a longer time scale, action prediction mechanisms may need to connect with the mentalization system (Frantz & Janoff-

Bulman, 2000; Malle & Pearce, 2001; Marsh et al., 2010) to allow inferences based on the goals and intentions of an actor. Future work will be needed to investigate the commonalities and differences among mechanisms involved in predicting actions over a range of time scales and in more complex social environment (e.g., interactions among multiple actors).

Chapter 5 Flash lag effect of skeleton posture action prediction

Imagine you walk by the busy corner of a train station. With a brief glance at the crowd, you can not only recognize which pedestrians are walking in leisure and which are running in a hurry, but also you can predict other people's actions in the immediate next moment (e.g., expecting someone to extend a hand to wave goodbye to a friend, or to pick up an object dropped on the ground) or in the near future (e.g., walking to the ticket office after stopping in front of the station information board). More generally, perception of biological motion—complex motion patterns of living organisms—not only enables efficient detection and recognition of actions in dynamic environments, but also supports more sophisticated abilities to predict future actions and make inferences about the intentions and goals of others.

While numerous studies have investigated recognition of biological motion (for a review, see Blake & Shiffrar, 2007) and identification of detailed attributes of actors involved in observed actions (Cutting & Kozlowski, 1977; Pollick, Lestou, Ryu, & Cho, 2002; van Boxtel & Lu, 2011), action prediction—the pinnacle of biological motion perception—has received much less attention. Early priming studies showed that observers automatically anticipate the future posture of observed actors, facilitating recognition of a test posture matching that which would be expected had the action continued (Kourtzi & Shiffrar, 1999; Verfaillie & Daems, 2002). Researchers also obtained evidence supporting a form of representational momentum for biological motion, as the final stopping position of an action as remembered by an observer was shifted forward in the direction of future movements (Jarraya, Amorim, & Bardy, 2005; Thornton & Hayes, 2004). Recent studies have shown that the visual system is able to predict future postures with very high temporal precision, implying a simulation process that operates almost in real-time to enable the prediction of future actions (Graf et al., 2007; Manera, Schouten, Verfaillie, & Becchio, 2013; Sparenberg, Springer, & Prinz, 2012; Springer & Prinz, 2010). In addition, trained experts such as CCTV (closed circuit television) operators show greater sensitivity of predicting situations that would end in violence than do novice observers (Troscianko et. al., 2004; Petrini et. al., 2014).

Previous research has shown that perception of biological motion is an active process, rather than a passive reconstruction of the visual input. When observing an action, participants automatically predict actions to project the future course of the activity. However, the mechanisms underlying action prediction still remain unclear. Previous work showed that the visual system can predict the whole body posture at a specific time point in the near future, which we term *global processing*. The existence of global mechanisms for predicting the future postures of body movements is highly plausible, since previous research on biological motion perception supports the existence of the mechanism for processing global posture information to recognize actions based on the change of body structure over time (Pinto & Shiffrar, 1999; Lu, 2010). For example, humans can recognize biological movements even when local motion information about joints is eliminated, as long as the form information provided by body posture is retained to a certain degree (Beintema & Lappe, 2002; Thurman & Lu, 2013).

We extended the *flash-lag* paradigm, to investigate the global mechanisms that predict the posture change resulting from whole-body movements—other than the local mechanisms that predict future positions of joints involved in action prediction. The FL effect, in which observers perceive that a briefly flashed object appears to lag behind a moving target even when they are in perfect alignment physically (Mackay, 1958; Nijhawan, 1994), has been primarily studied for movements of simple objects (e.g., a moving bar, a rotating dot) (Durant & Johnston, 2004; Moore & Enns, 2004; Whitney, Murakami, & Cavanagh, 2000; Nijhawan, 2002; Whitney, 2002). To our knowledge, the FL effect has not been examined for complex movement patterns such as biological motion. The present study uses the FL effect as a means to measure human ability to predict future actions. We assessed whether the FL effect is primarily driven by action prediction based on sophisticated representations with a hierarchical movement structure. In addition, we try to use this paradigm to dissociate local and global mechanisms that may underlie action prediction. To examine the global mechanism for predicting the whole-body posture in the future, a stationary posture was briefly flashed during the viewing of a walking action, measuring the motion-induced posture change, reminiscent of representational momentum studies in the literature on action prediction (Jarraya, Amorim, & Bardy, 2005; Thornton & Hayes, 2004). We expected that observers are more likely to judge that the walking posture lagged behind the flash, even though it is physically aligned with the flash. Taken together, our findings would provide psychophysical evidence that biological motion perception is predictive. By updating an internal

model, people are able to anticipate the future position of critical joints and the future posture of whole-body movements.

5.1 Experiment 1: Posture flash lag effect

The previous experiments reported in Chapter 4 showed that human observers employ local mechanisms for predicting joint positions in the immediate future. Experiment 1 aimed to examine the global mechanism for predicting the whole-body posture in the immediate future. If action prediction also automatically employs the global mechanism, we would expect to discover a FL by flashing a stationary posture during the viewing of an action.

5.1.1 Methods

5.1.1.1 Participants

Thirty-nine UCLA undergraduate students (25 females, 14 males, M_{age}= 21.03). Twenty of them completed upright block first and 19 of them completed inverted block first. They participated in this one-hour experiment to receive one class credit. Participants reported normal or corrected-to-normal vision.

5.1.1.2 Design and Stimuli

Experiment 1 involved the same walking action as in the previous experiments, but stimuli were presented in a limited life-time display to encourage global processing in biological motion perception (Beintema & Lappe, 2002). As shown in Figure 5-1, we randomly sampled points along the limbs to yield a walker with 17 white dots, including one head dot and two dots on each of the eight limbs. Each dot had a lifetime of ten frames (i.e., 166.7 ms). The limited lifetime display removes the smooth motion trajectories of joints, thereby reducing contributions from local motion prediction, allowing us to focus on effects attributable to global mechanisms based on posture changes over time in action processing.

On each trial, at some point during the second step in the walking cycle, a line-drawing skeleton in the color of gray with a line width of 0.07 visual degrees was flashed for 33 ms. The flash skeleton always appeared at the same physical location and with the same posture across all the conditions, but at different time points. The flash skeleton was placed at the same reference frame used in previous experiments (i.e., frame 53). As shown in Figure 5-1, the flash skeleton appeared several frames prior to, in the same frame as, or several frames after the frame in which

the moving walker with limited-lifetime dots perfectly aligned with the flash; these stimuli were respectively labeled *behind*, *aligned*, and *ahead*, based upon the relative temporal relation between the moving walker and the stationary posture at the time of the flash. Similar to the previous experiments, the first within-subject factor was body orientation of the red dotted walker (upright vs. inverted), and each participant completed two blocks of experiment. The second within-subject independent variable was the temporal offset level. The experiment had seven temporal offset levels varied from -100 to +100ms with a step size of 33.3ms. The temporal offset level was randomized within each block. There were 40 trials per temporal offset level such that there were 280 trials in each block. The experiment also had between-subject counterbalancing factor, the order of blocks (upright block first or inverted block first). The walker always faced left and walked forward on treadmill in all trials.
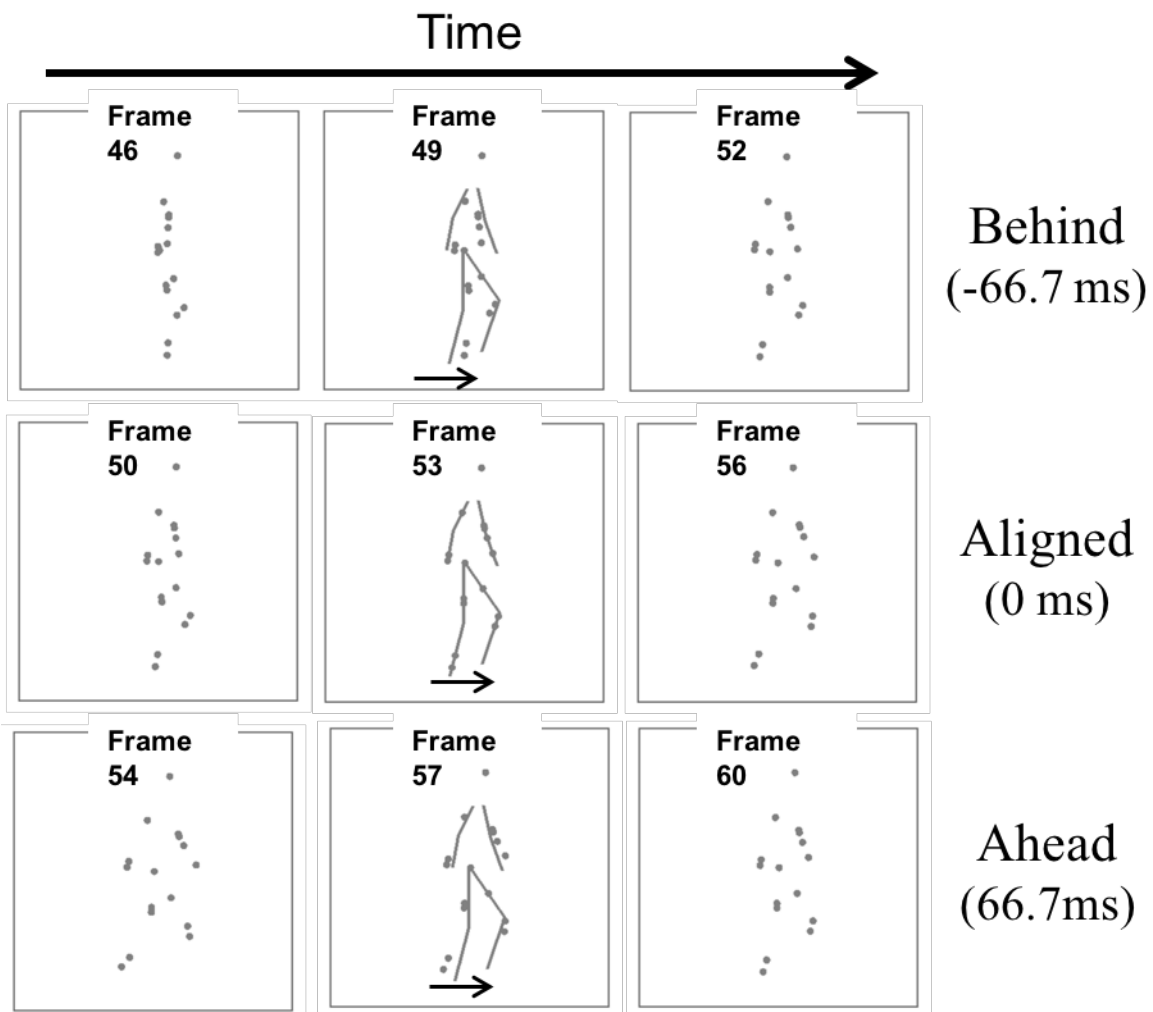
Figure 5-1. Illustration of the stimulus used in Experiment 2. The walker consisted of limited-lifetime dots that were physically behind, aligned with or ahead of the stationary flashed skeleton. The arrow indicates the facing direction of the walker.

5.1.1.3 Procedure

Each trial started with a central fixation cross, followed by the walker consisting of limited-lifetime dots. A posture in the skeleton display was briefly flashed at different time points depending on the temporal offset between the flash and the walker. Participants were asked to judge whether the moving walker appeared aligned with the flash skeleton (yes or no). Participants were allowed to indicate that they did not observe the flash. The missing trials were removed from the analysis.

Similarly, participants completed 10 practice trials prior to testing. The presentation of the practice trials was twice slower than that in the experimental trials. The action duration lasted 2.03s in the practice session, which doubled the duration relative to the experimental trials. The experimental session consisted of two blocks, each including 140 upright trials or 140 inverted trials. The order of the two blocks was counterbalanced across subjects. Within each block, the order of the offsets was randomized.

5.1.2 Results

We removed all the trials with "missing" responses (3.3% of trials across all participants) from data analyses. Figure 5-2 shows the mean proportion of responses in which observers reported the walker as aligned with the flash skeleton, as a function of temporal offsets between the walker and the flash skeleton. If observers perceived veridical postures of the walker without action prediction, the highest "aligned" response proportion would be reported when the flash skeleton and the walker was physically in perfect alignment (i.e., zero offset). However, for the upright walker, we found that observers were more likely to judge the walker that lagged 33 ms behind the flash as aligned with the flash skeleton, compared to the zero offset in which the walker was physically in alignment with the flash (when offset was zero) ($t(1,38)=2.21$, $p = .033$). Hence, the peak of the function in Figure 5-2 was shifted away from the zero offset and towards a negative offset. This shifted peak reveals the existence of FL in terms of global posture change. This posture FL effect depended on the body orientation of the walker, as the effect size was much smaller for the inverted condition, in which the "aligned" response proportion with an

offset lag of 33ms was the same as when the walker was physically in alignment with the flash (t(38) = .904, p=.372).
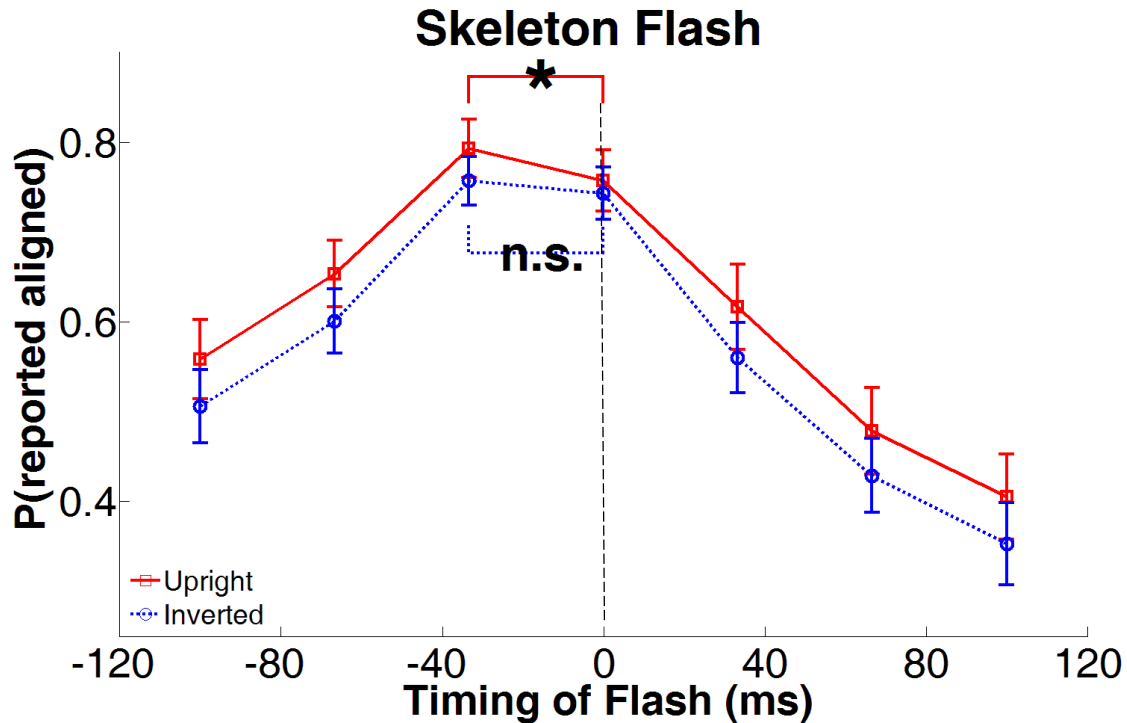
Figure 5-2. Results for posture FL effect in Experiment 2. The mean proportion of responses that observers reported the walker in alignment with the flash skeleton is shown as a function of temporal offsets between the walker and the flash skeleton. Zero offset indicates physically perfect alignment between the flash skeleton and the walker. Negative (positive) offsets indicate that the walker physically lagged behind (ahead of) the flash skeleton.

To examine the directionality of the shifted offset when observers reported alignment between the walker and the flash, we conducted a repeated-measures ANOVA with three within-subjects factors, 2 (orientation: upright vs. inverted) × 2 (offset directionality: behind vs. ahead of)× 3 (offset magnitude: 2, 4, 6 frames). This analysis did not include the condition of zero offset. We found higher response proportions for the negative offset than for the positive offset ($F(1,38)$=31.86, p <.001, η_p^2= .46), suggesting participants were more likely to judge that the walker that moved with lags was in alignment with the flash posture, than the walker that moved ahead of the flash. This result indicates a "predictive shift" in perceiving a moving walker in relation to its future postures. The analysis also revealed a significant main effect of body

orientation, as participants were more likely to perceive the flash skeleton in alignment with an upright walker than with an inverted walker ($F(1,38)=5.60$, $p = .023$, $\eta_p^2= .13$). This result is consistent with the impact of body orientation on the joint FL effect found in the first two experiments, and confirms that the posture FL effect was primarily driven by action processing, given the essential role of natural body orientation in biological motion perception. Finally, there was a significant main effect of the magnitude of the offset between the flash and the walker ($F(1,38)=51.62$, $p <.001$, $\eta_p^2= .58$), suggesting that people were sensitive to the relation between the flash skeleton relative to the moving walker. No any significant interaction effects were obtained in this analysis.

We then analyzed data separately based on the order of the blocks. An order effect was detected when we analyzed the effect of body orientation. For the upright walkers, observers were more likely (approaching to significance) to judge the walker that lagged 33 ms behind the flash as aligned with the flash skeleton, compared to the zero offset in which the walker was physically in alignment with the flash (when offset was zero) (.807 vs. .767, $t(19)=1.82$, $p = .084$) when participants viewed upright walker in the first block (red line in Figure 5-3). Such difference (approaching to significance) was also found when participants viewed inverted block first (.787 vs. .753, $t(18) =1.74$, $p=.099$, as redline Figure 5-4). However, for inverted walker, observers showed no difference between -33ms and 0ms conditions ($t(19)=0.86$, $p = .400$) when participants viewed upright walker in the first block (blue line in Figure 5-3) but showed significant difference (.748 vs. .698, $t(1,18)=2.52$, $p = .021$) when participants viewed inverted walker in the first block (blue line in Figure 5-4). This results indicated that viewing order modulated the global representation of walking action. Upright walker's 'predictive shift' to the future (i.e. flash skeleton 'shift' to the past) is relatively robust regardless of the viewing order (i.e. flash lag effect in either first block or second block for upright walker). However, viewing upright walker first reduced the global processing efficiency of inverted walker in the second block. Vice versa, in contrast, viewing inverted walker first exerts no influence on global processing of the upright walker which was view in the second block.
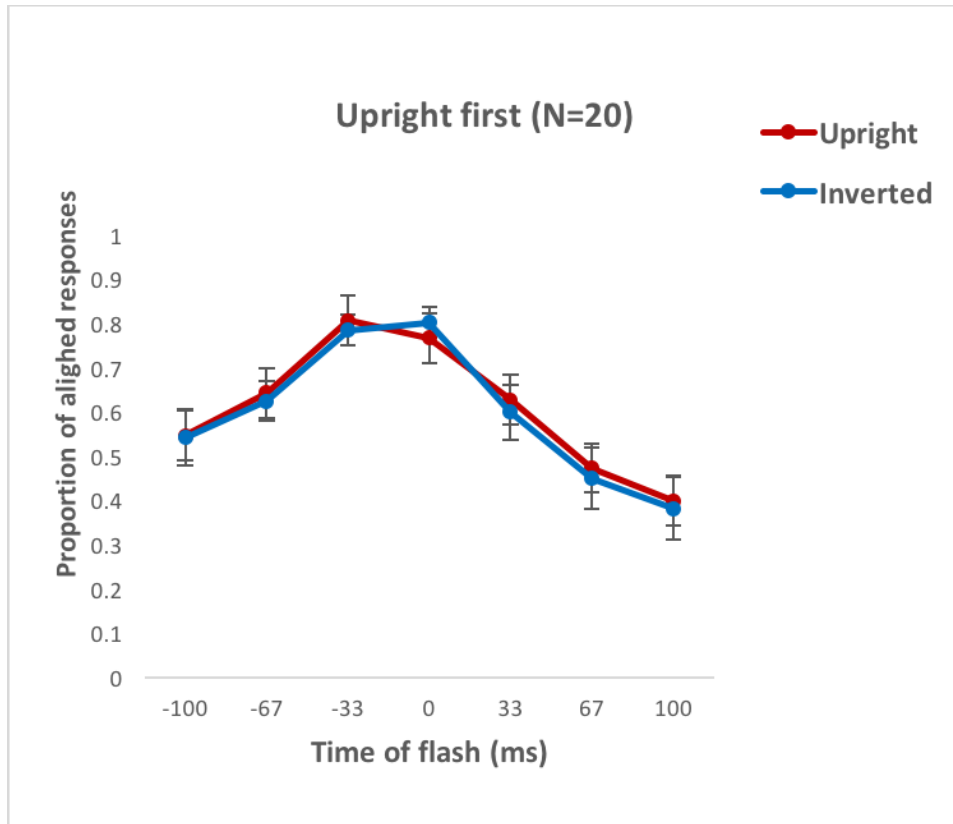
Figure 5-3. Results for the posture flash-lag effect in upright first experiment (N=20).
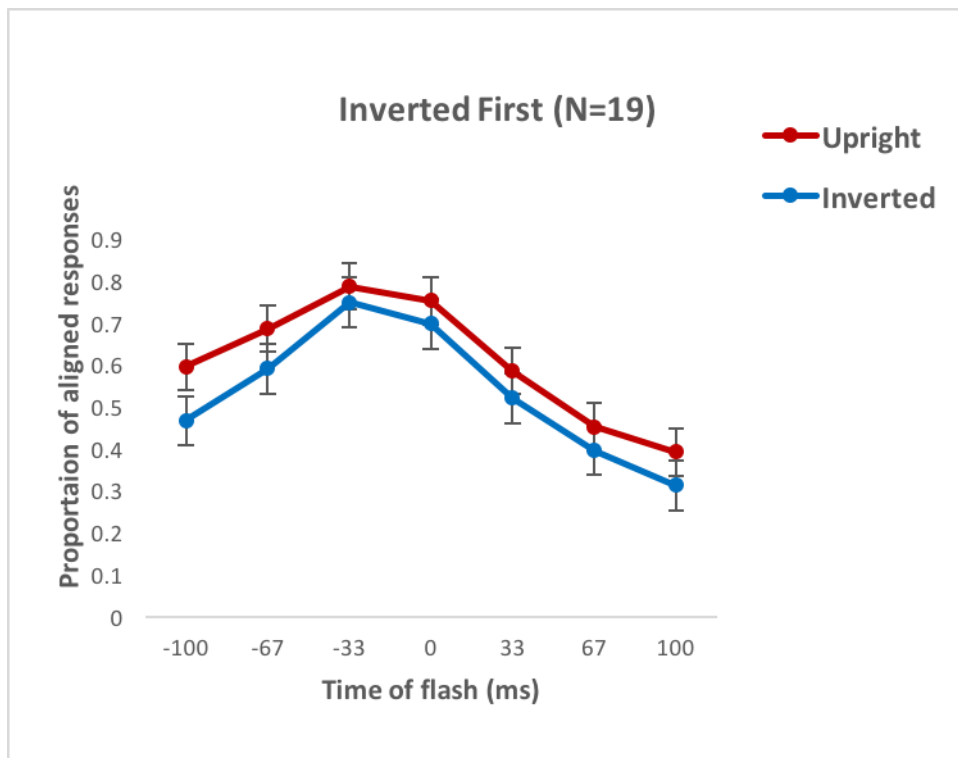
Figure 5-4. Results for the posture flash-lag effect in inverted first experiment (N=19).

5.2 Experiment 2: Replication with different parameters of display

As mentioned in Chapter 1, the perception of biological motion should be regarded as a phenomenon with multiple levels, including at least two distinct underlying mechanisms: one is at the local level that exploits the local motion of the individual dots or body parts (such as joints and limbs). The other one is at the global level that retrieves structural information (specifically, the global and dynamically changing shape of a body in motion).

Our study aimed to investigate flash-lag effect for human action at both the local level and the global level. We also examined how body orientation influenced the strength of flash-lag effect. In the previous chapter, we conducted joint flash-lag effect experiments. We found flash-lag effect for the local image motion. When the local image motion was familiar to the observers (when the walker walked forward and the local image motion was from foot joint), upright walking motion induced stronger flash-lag effect. In the current experiment, we investigated posture flash-lag effect. Again, we expected a flash-lag effect by flashing a stationary posture during the walking cycle of a limited lifetime display. And we predicted that body orientation would modulate the degree of misperception, such that the motion from upright walker would induced stronger misperception. To exclude low-level visual feature influence, we plan to replicate Experiment 1 with different parameters (i.e. color of stimuli; number of dots per frame; life time of dots)

5.2.1 Methods

5.2.1.1 Participants

Forty-four UCLA undergraduate students (8 males, M_{age}= 20.51) participated in this one-hour experiment to receive one class credit. Participants reported normal or corrected-to-normal vision. Fifteen of them participated two-block experiment (8 of them completed upright block first and 7 of them completed inverted block first). Twenty-nine of them participated one-block experiment (15 of them completed upright block and 14 of them completed inverted block).

5.2.1.2 Design and Stimuli

We used the same walker as Experiment 1 with same design except a few changes of stimulus parameters described below. The limited life-time display of the walker was made up by nine red dots, to further prompt the involving of global processing in biological motion perception (Beintema & Lappe, 2002). The limited life-time display (see Figure 5-5) included one head dot and one dot on each of the eight limbs (in contrast to 2 dots per segment in Experiment 1). Eight dots on limbs were randomly sampled along the limbs. These eight dots had a 'limited lifetime' of 5 frames (i.e., 83.35ms), after which it disappeared and was redrawn in another randomly sampled position along the limb. The initial lifetime was randomly assigned for each dot to make the dots disappear at different time points. Such limited lifetime display made tracking local motion from joints difficult so that the contributions from local motion prediction were reduced.
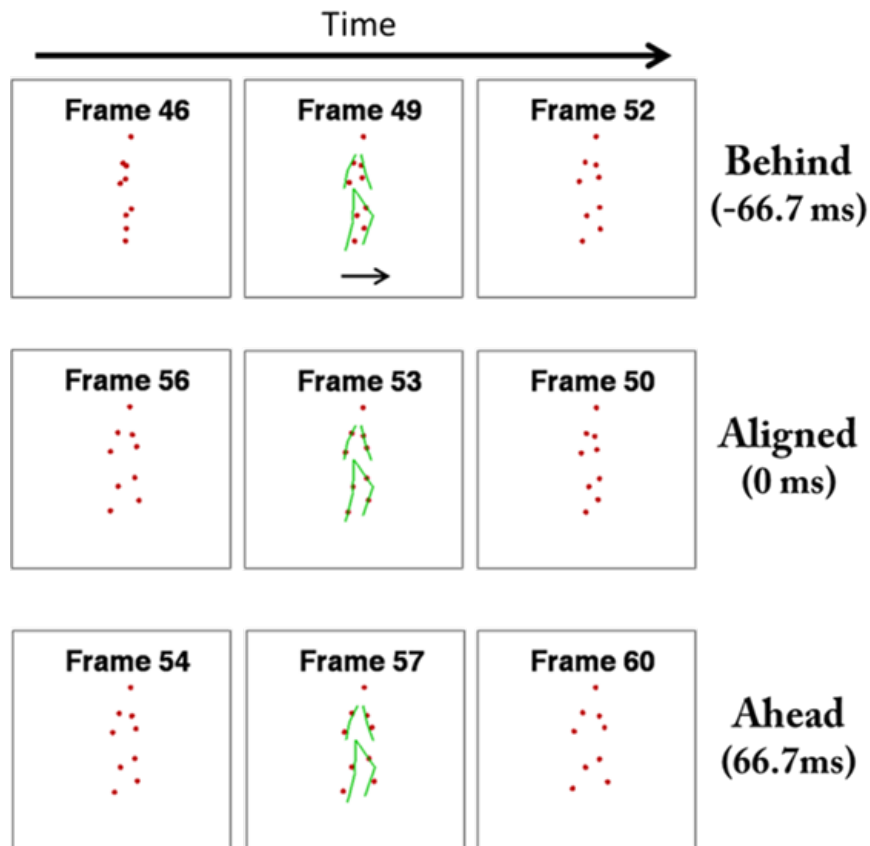
Figure 5-5. Illustration of the stimulus used in Experiment 2. The walker consisted of limited-lifetime dots that were physically behind, aligned with or ahead of the stationary flashed skeleton. The arrow indicates the facing direction of the walker.

5.2.1.3 Procedure

Similar to Experiment 1, before each trial started, a white cross-fixation appeared in the center of the screen shortly to help participants to pay attention to the center of the screen. When the cross-fixation disappear, the red dotted walker was shown in the center of the screen and completed one walking cycle which included 60 frames and lasted one second. A green skeleton with eight limbs was then briefly flashed for 33.3ms during the waking cycle of the red dotted walker. The green skeleton appeared at the same physical location with same posture across all conditions, but the time points of flashing it varied (the skeleton was shown at different frames) across trial. Thus, the green skeleton would be physically *behind*, *aligned with*, or *ahead of* the red dotted walker (see Figure 5-5). At the end of each trial, participants were asked to judge whether the green skeleton was aligned or misaligned with the red dotted walker, or they could indicate that they missed the trial. The flash skeleton was misaligned with the red dotted walker in six out of seven temporal offset levels thus the strength of misperception was measured in terms of the proportion of alignment responses.

5.2.2 Results

Similar to Experiment 1, all the trials with the "missing" responses were removed from data analyses. Figure 5-6 shows the mean proportion of responses in which observers reported the walker as aligned with the flash skeleton, as a function of temporal offsets between the walker and the flash skeleton. If observers perceived veridical postures of the walker without experiencing flash lag effect, the highest "aligned" response proportion would be reported when the flash skeleton and the walker was physically in perfect alignment (i.e., zero offset). However, for the upright walker, we found that observers were more likely to judge the walker that lagged 33 ms behind the flash as aligned with the flash skeleton, compared to the zero offset in which the walker was physically in alignment with the flash (when offset was zero) ($t(14)=2.93$, $p = .011$). Hence, the peak of the function in Figure 5-6 was shifted away from the zero offset and towards a negative offset. The mean proportion of aligned responses was highest at -33.3ms temporal offset level regardless body orientation. When the walker was inverted, the mean

proportion of aligned response at -33.3ms temporal offset level was also significant higher than at 0ms temporal offset level (t (14) =2.55, p=.023).

We then analyzed data separately based on the order of the blocks (Figure 5-7: upright first vs. Figure 5-8: inverted first). An order effect was detected when we analyzed the effect of body orientation. Upright walker had higher mean proportion of aligned responses than inverted walker when the order was upright first at -33.3ms temporal offset level (.684 vs. .595, t (7) =2.35, p=.05). But such difference was not found when participants ran inverted block first (.665 vs. .653, t (6) =0.022, p=.98).
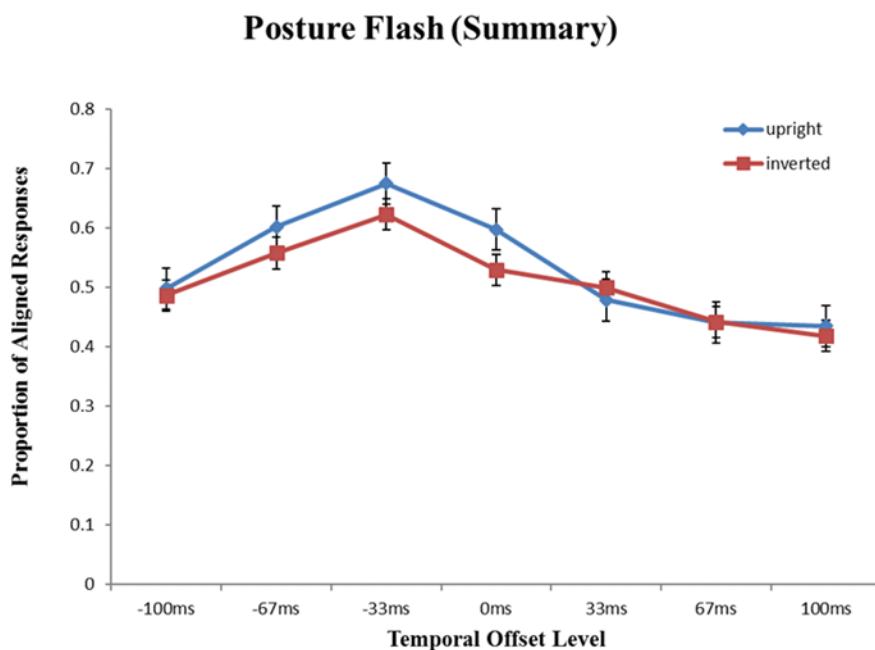
Figure 5-6. Results for the posture flash-lag effect in two-block experiment (N=15). The mean proportion of responses that observers reported the walker in alignment with the flash skeleton is shown as a function of temporal offsets between the walker and the flash skeleton. Zero offset indicates physically alignment between the flash skeleton and the walker. Negative (positive) offsets indicate that the walker physically lagged behind the flash skeleton (flash ahead of condition).
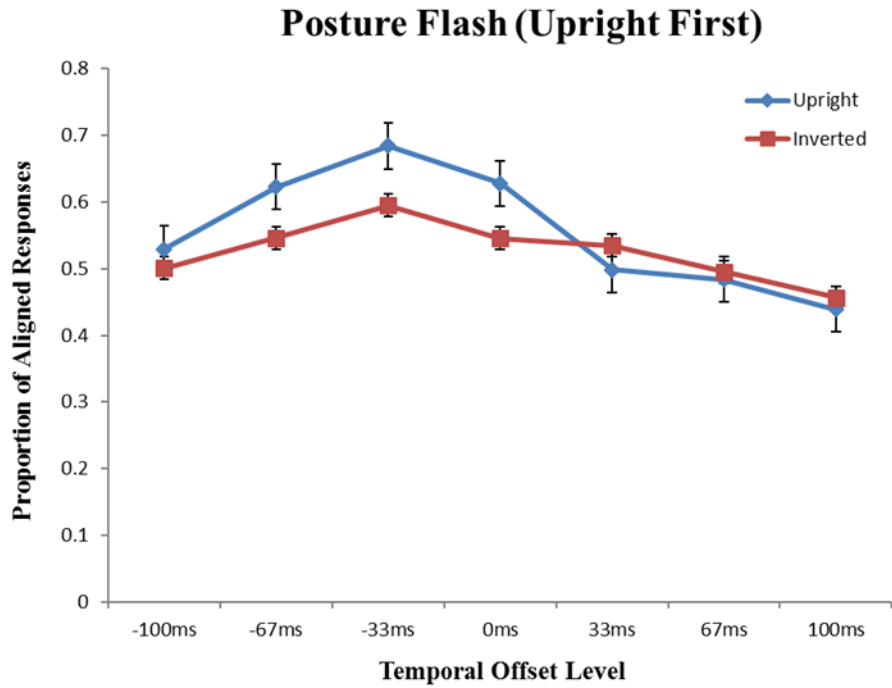
Figure 5-7. Results for the posture flash-lag effect in upright first two-block experiment (N=8).
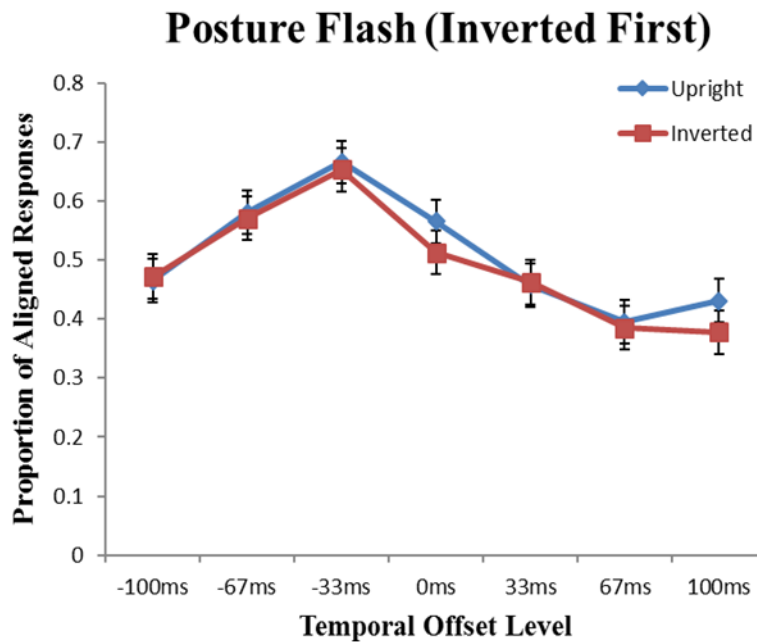
Figure 5-8. Results for the posture flash-lag effect in inverted first two-block experiment (N=7).

Then we analyzed the data from first block only to reduce the influence of the possible order effect. We took the first block data from the two-block experiment. In order to run an independent sample t-test, we had another 29 participants completed one-block experiment to increase the sample size. At -33.3ms temporal offset level, the upright walker had higher mean proportion of aligned responses than inverted walker. Such difference approached significance level, $t(42) = 1.987$, $p=.05$.

Again, to examine the directionality of the shifted offset when observers reported alignment between the walker and the flash, we conducted a repeated-measures ANOVA with three within-subjects factors, 2 (orientation: upright vs. inverted) × 2 (offset directionality: behind vs. ahead of)× 3 (offset magnitude: 2, 4, 6 frames). This analysis did not include the condition of zero offset. We found higher response proportions for the negative offset than for the positive offset ($F(1,42)=12.19$, $p <.001$, $\eta_p^2= .23$), suggesting participants were more likely to judge that the walker that moved with lags was in alignment with the flash posture, than the walker that moved ahead of the flash. This result indicates a "predictive shift" in perceiving a moving walker in relation to its future postures. Also, there was a significant main effect of the magnitude of the offset between the flash and the walker ($F(2,84)=13.75$, $p <.001$, $\eta_p^2= .25$), suggesting that people were sensitive to the relation between the flash skeleton relative to the moving walker. A significant 3-way interaction effect was obtained in this analysis ($F(2,84)=13.75$, $p <.001$, $\eta_p^2= .25$), indicating the directionality (i.e. future vs. past) modulation on offset is dependent on body orientation. For upright body configuration, the 2-way ANOVA showed significant interaction of directionality and offset ($F(2,42)=15.47$, $p <.001$, $\eta_p^2= .42$). However, the modulation is not significant for inverted body configuration. The results demonstrated that the "predictive shift" to the future posture (i.e. asymmetry of future and past) in action perception is modulated by body configuration, or action representation.

5.3 General Discussion

The results of the present studies and previous chapter demonstrated that biological motion perception is predictive, based both on local mechanisms for predicting joint position, and global mechanisms for predicting the whole-body posture in the near future. These two mechanisms were investigated using the flash-lag paradigm. Previous experiments measured the FL effect on local joint movements by focusing on a critical joint in walking action (foot), and an important

but less critical joint (hand). An FL effect was obtained in several different conditions, supporting the hypothesis that prediction for dynamic stimuli is a primitive operation for the visual system. However, the magnitude of the FL was influenced substantially by the characteristics of the action. In particular, observing a walker with the natural body orientation (i.e., upright) yielded significantly stronger FL for the critical joint of foot than viewing an inverted walker. This difference cannot be attributed to general motion extrapolation based on low-level motion cues (e.g., velocity, direction, or acceleration), since the characteristics of foot movement trajectories in the upright and inverted conditions were well matched. A control experiment provided converging evidence that the impact of body orientation on the FL is a signature of action prediction.

This study demonstrated the existence of a global mechanism for anticipating changes in whole-body posture over time. The results showed that people were able to recognize human motion without local image motion and we found flash-lag effect in global human motion. When a skeleton posture of an upright walker was flashed briefly, observers were more likely to judge the flash as aligned with the walker posture 33ms ahead of the "current" movement in the future, rather than the posture physically aligned with the flash "right now". At the -33.3ms temporal offset level, the green skeleton represented an immediate future posture of the red dotted walker. Perhaps observers extrapolated the trajectory of the moving red-dotted walker into the future and thus perceived that the walker aligned with the flash skeleton (Nijhawan, 1994). However, such results can alternatively be explained by the post-diction theory (Eagleman & Sejnowski, 2000). After the flash, the walker immediately reached its future posture which was the green skeleton represented. The observers integrated the post-flash motion and perceived that two stimuli were aligned. But neither motion extrapolation or post-diction theory alone can fully explain the impact of body orientation on the flash-lag effect in biological motion. We have natural preference for upright stimulus (Simion et al., 2008) due to more experience with upright action observing. Therefore, we are more familiar with upright motion and probably have better templates for them to recognize upright motio. The body orientation modulated the degree of misperception. In the one-block experiment analysis, we found that the upright walker had higher mean proportion of aligned responses than inverted walker at -33.3ms temporal offset level, and this difference approached significance level. This suggested the orientation specificity

of biological motion perception, again indicating the involvement of action prediction processing.

The existence of a posture FL effect agrees with previous findings that action prediction employs real-time simulation of change in body posture to accurately predict an action sequence interrupted by a temporal occluder (Graf et al., 2007). The posture FL effect was weakened in the inverted condition. Furthermore, limited lifetime postures that temporally lagged behind the flash skeleton in the walking action (i.e. -100ms, -67ms, and -33ms) were perceived as more aligned with the flash than postures aligned or ahead of the flash skeleton (i.e. 0ms, 33ms, 67ms, and 100ms), suggesting that action prediction employs the global mechanism to predict the change of the whole-body posture in the near future. Although inversion of body orientation weakens global processing in biological motion perception, it does not eliminate the use of the global mechanism in predicting future posture even when the action is presented upside-down. This finding is consistent with previous studies of the inversion effect in biological motion perception (Pavlova & Sokolov, 2000; Sumi, 1984), and highlights the robust use of the global mechanism in predicting future actions.

An order effect of the block design was discovered, meaning that viewing experience altered observers' top-down modulation of body configuration on FL effect. When viewing the upright walker in the first block, observers reported more aligned responses in -33ms condition (i.e. when the flashed skeleton was shown as 33ms in the future) in the upright condition compared to the inverted condition, indicating the involvement of action prediction processing in posture FL effect task performance. However, when viewing the inverted walker first, the difference between upright and inverted conditions disappeared, showing that lack of intact body representation during previous immediate viewing experience may lead to impaired template comparison (Lange, Georg, & Lappe, 2006; Lange & Lappe, 2006) or reduced global processing in action prediction.

Chapter 6 Conclusion and Future Questions

In this thesis, I summarized four series of studies to examine biological motion perception. These range of experiments touched on three important aspects including action recognition, action prediction and action interaction inferences.

First, we examined whether social interactive activities, as opposed to solitary actions from a single individual (e.g. walking), impact the rivalry between two competing action stimuli to determine what action information reaches awareness and what information does not. We found that actions eliciting natural body movements and meaningful interpersonal activities are granted preferential access to visual awareness.

Second, we investigated how action prediction performance depends on the involvement of interpersonal interactions, by comparing solitary actions with a single actor versus interpersonal actions with social interactions. We found that people can do real-time action prediction and give a good estimation of the future action even after 400ms. In the absence of an interactive actor, people did better in precise online temporal estimation ('when' or predictive timing), however, when there is a second actor interacting with the test actor, the prediction showed large uncertainty in temporal dimension, but became more accurate in judging spatial action sequences and also in detecting rotation of the original action trajectory even when the postures don't match perfectly in temporal domain ('what' or predictive identification).

Third, we investigated whether action representations interact with generic motion mechanisms underlying position processing of dynamic stimuli to evoke the misperception of joint location in a moving body. We found robust flash-lag effects for localizing joints in a walking action, supporting the hypothesis that mislocalizing spatial positions of moving objects in dynamic stimuli results from a primitive operation of the visual system, which can be simulated by temporal averaging model (e.g. the presence and absence of a flash-lag effect, and changes in the size of the effect across different joints— basically due the faster movements in foot joints than hand joints). However, the magnitude of the FL was influenced substantially by the characteristics of the action (e.g. action representation, body configuration, and familiarity of actions), which cannot be explained by the model with generic mechanism for motion

perception. This result reveal the important role of action representation and prediction in localizing moving body when humans perform actions.

In the last Chapter, we investigated whether people were able to apply human motion prediction without local image motion. We found biological motion perception is predictive, based both on local mechanisms for predicting joint position, and global mechanisms for predicting the whole-body posture in the near future. This study demonstrated the existence of a global mechanism for anticipating changes in whole-body posture over time. Although inversion of body orientation weakens global processing in biological motion perception, it does not eliminate the use of the global mechanism in predicting future posture even when the action is presented upside-down. This finding is consistent with previous studies of the inversion effect in biological motion perception (Pavlova & Sokolov, 2000; Sumi, 1984), and highlights the robust use of the global mechanism in predicting future actions.

Although my thesis work provides candidate answers to some of important questions in the field of biological motion perception, many questions remain open.

For action prediction, given the evidence of temporal precision when people predict actor posture after certain temporal or spatial occlusion, a theory of "online action simulation" is proposed. The possible underlying neural mechanisms can be increased neural firing rates biased toward expectation or enhanced synchronization when anticipating the signal. However, it is still largely unknown how this neural modulation is related to predicting various actions of different movement styles and speeds. My thesis showed that online action simulation may not be necessarily if temporal average model is sufficient to explain the results under modulation from action representation. Especially, when the post flash action is missed (i.e. Experiment 4.3), the perceived position of the moving limb is biased toward the past rather than the future position, indicating the failure of online simulation to counteract or cancel out the temporal averaging mechanism. More cognitive neuroscience research can be conducted to address this point.

For action interaction inference, evidence shows that people can utilize the correspondent interactive information to enhance stimulus visibility in noisy background. Also, humans can infer mental states of non-figure-like symbols based on their animated movement paths. However, the underlying computational algorithm used by the brain to do such inference is still an open question. One possible model underlying intention/causal inference may be the Bayesian

empirical inference, which adjusted predictive errors based on the generative model at each level from kinematics (STS) to intention/goal representation (parietal and frontal MNS). Another unanswered question is which aspect of action interactive information is used by the brain to enhance biomotion perception of any single actor? Is it the general action type or movement style or overall synchronization of movement tempos, or the more specific alignment of body postures during certain moments? Does it relate to the action understanding ability of individual subjects? Would this effect hold if subjects interpret interactive actions differently? These are intriguing questions to address in the future research about action inference mechanisms.

Development studies showed that the most amazing results is 2-day-old infants' preference to biomotion stimuli compared to general motion patterns. This preference is not limited to human movement but also other animals' motion. These findings push biomotion perception toward the nature/innate end, rather than nurture/learned perceptual capability. If biologically wired wrongly, Autistic Spectrum Disorder (ASD) children will exhibit social cognition deficits in early years, which lead to impairments in their thinking, feeling, language, and the ability to relate to others in later years all the way till adulthood. One possible area involved would again be STS, which is not only associated with impaired biomotion perception, but also ability to recognize gaze direction and mouth movement related to social facial communication. Mirror Neuron System (MNS) can also be a candidate region if it is involved in the common coding processes in biomotion perception. Indeed, some autistic patients show motor deficits as well, especially the lack of the flexibility and over rigidity of motor responses. Other patients (e.g., schizophrenia and Obsessive-Compulsive Disorder) also suffer similar problems. The coincidence of their abnormal social cognition and impaired biomotion perception may be due to a decoupling of STS activity from other areas of the social brain network. Moreover, patients with schizophrenia seem to present unrealistic alterations of the perception of specific social stimuli, such as face emotion and biological motions. These alterations include the typical symptoms of over-interpreting stimuli as social and intentional, which is expressed in paranoid symptoms and persecutory delusions. Part of the results I did not report here is the correlation between inversion modulation on actor visibility in binocular rivalry (i.e. Experiment 2.1) and autistic quotient (AQ) scores. We found that among normal population, the higher the autistic traits, the smaller the body inversion effect on action visibility. The correlation between biomotion perception and AQ has been replicated in other but not all studies in our lab,

providing opportunities for further studies to disentangle the complexity of multi-component or multi-level biomotion mechanism. Thus, the influence of high AQ on each component or level of biomotion processing may vary accordingly.

Finally, biomotion perception can be enhanced substantially through motor training on top of visual experience, again indicating the validity of "common coding theory". It is also modulated by adaptation, attention, task instructions or sensory modalities. The specific sensitivity to biomotion compared to general motion patterns is robust across human ages, culture or social context and even species, indicating the evolutionary significance of this ability for animals' survival. In the future, it would be intriguing to test biomotion inference among different groups of participants with broad range motor expertise or training experiences, to contribute to the nature/nurture debate in cognitive psychology.

Reference

Ahlstrom, V., Blake, R., & Ahlstrom, U. (1997). Perception of biological motion. *Perception, 26*(12), 1539-1548.

Alais, D., & Blake, R. (1999). Grouping visual features during binocular rivalry. *Vision Research, 39*(26), 4341-4353.

Alais, D., & Blake, R. (2005). *Binocular rivalry*. Cambridge, Mass.: MIT Press.

Albright, T. D., & Desimone, R. (1987). Local Precision of Visuotopic Organization in the Middle Temporal Area (Mt) of the Macaque. *Experimental Brain Research, 65*(3), 582-592.

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends Cogn Sci, 4*(7), 267-278.

Ambady, N., Hallahan, M., & Conner, B. (1999). Accuracy of judgments of sexual orientation from thin slices of behavior. *Journal of Personality and Social Psychology, 77*(3), 538-547.

Andreou, L. V., Kashino, M., & Chait, M. (2011). The role of temporal regularity in auditory segregation. *Hearing Research, 280*(1-2), 228-235. doi:DOI 10.1016/j.heares.2011.06.001

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception, 33*(6), 717-746. doi:Doi 10.1068/P5096

Baldo, M. V. C., & Klein, S. A. (1995). Extrapolation or Attention Shift. *Nature, 378*(6557), 565-566.

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions (vol 11, pg 280, 2007). *Trends in Cognitive Sciences, 11*(9), 372-372. doi:DOI 10.1016/j.tics.2007.08.004

Barclay, C. D., Cutting, J. E., & Kozlowski, L. T. (1978). Temporal and Spatial Factors in Gait Perception That Influence Gender Recognition. *Perception & Psychophysics, 23*(2), 145-152.

Bardi, L., Regolin, L., & Simion, F. (2011). Biological motion preference in humans at birth: role of dynamic and configural properties. *Dev Sci, 14*(2), 353-359.

Bargh, J. A., & Williams, E. L. (2006). The Automaticity of Social Life. *Curr Dir Psychol Sci, 15*(1), 1-4.

Battelli, L., Pascual-Leone, A., & Cavanagh, P. (2007). The 'when' pathway of the right parietal lobe. *Trends in Cognitive Sciences, 11*(5), 204-210. doi:DOI 10.1016/j.tics.2007.03.001

Beintema, J. A., Georg, K., & Lappe, M. (2006). Perception of biological motion from limited-lifetime stimuli. *Perception & Psychophysics, 68*(4), 613-624. doi:Doi 10.3758/Bf03208763

Beintema, J. A., & Lappe, M. (2002). Perception of biological motion without local image motion. *Proceedings of the National Academy of Sciences of the United States of America, 99*(8), 5661-5663. doi:DOI 10.1073/pnas.082483699

Bertamini, M. (1993). Memory for Position and Dynamic Representations. *Memory & Cognition, 21*(4), 449-457.

Bertenthal, B. I., & Pinto, J. (1994). Global Processing of Biological Motions. *Psychological Science, 5*(4), 221-225.

Bertenthal, B. I., Proffitt, D. R., & Cutting, J. E. (1984). Infant Sensitivity to Figural Coherence in Biomechanical Motions. *Journal of Experimental Child Psychology, 37*(2), 213-230.

Bidet-Caulet, A., Voisin, J., Bertrand, O., & Fonlupt, P. (2005). Listening to a walking human activates the temporal biological motion area. *Neuroimage, 28*(1), 132-139. doi:DOI 10.1016/j.neuroimage.2005.06.018

Blake, R. (1977). Threshold conditions for binocular rivalry. *J Exp Psychol Hum Percept Perform, 3*(2), 251-257.

Blake, R. (1989). A neural theory of binocular rivalry. *Psychol Rev, 96*(1), 145-167.

Blake, R., & Logothetis, N. (2002). Visual competition. *Nat Rev Neurosci, 3*(1), 13-21. doi:10.1038/nrn701

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology, 58*, 47-73. doi:DOI 10.1146/annurev.psych.57.102904.190152

Blake, R., Yu, K., Lokey, M., & Norman, H. (1998). Binocular rivalry and motion perception. *J Cogn Neurosci, 10*(1), 46-60.

Blakemore, S. J., & Frith, C. (2005). The role of motor contagion in the prediction of action. *Neuropsychologia, 43*(2), 260-267. doi:DOI 10.1016/j.neuropsychologia.2004.11.012

Boussaoud, D. (2001). Attention versus intention in the primate premotor cortex. *Neuroimage, 14*(1), S40-S45. doi:DOI 10.1006/nimg.2001.0818

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis, 10*(4), 433-436.

Brascamp, J., Sohn, H., Lee, S. H., & Blake, R. (2013). A monocular contribution to stimulus rivalry. *Proc Natl Acad Sci U S A, 110*(21), 8337-8344. doi:10.1073/pnas.1305393110

Brown, W. M., Cronk, L., Grochow, K., Jacobson, A., Liu, C. K., Popovic, Z., & Trivers, R. (2005). Dance reveals symmetry especially in young men. *Nature, 438*(7071), 1148-1150. doi:Doi 10.1038/Nature04344

Brownlow, S., Dixon, A. R., Egbert, C. A., & Radcliffe, R. D. (1997). Perception of movement and dancer characteristics from point-light displays of dance. *Psychological Record, 47*(3), 411-421.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., V, G., . . . Freund, H. J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience, 13*(2), 400-404. doi:DOI 10.1046/j.1460-9568.2001.01385.x

Bulthoff, I., Bulthoff, H., & Sinha, P. (1998). Top-down influences on stereoscopic depth-perception. *Nat Neurosci, 1*(3), 254-257. doi:10.1038/699

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology-Human Perception and Performance, 27*(1), 115-127.

Centelles, L., Assaiante, C., Nazarian, B., Anton, J. L., & Schmitz, C. (2011). Recruitment of both the mirror and the mentalizing networks when observing social interactions depicted by point-lights: a neuroimaging study. *Plos One, 6*(1), e15749. doi:10.1371/journal.pone.0015749

Chang, D. H. F., & Troje, N. F. (2008). Perception of animacy and direction from local biological motion signals. *Journal of Vision, 8*(5). doi:10.1167/8.5.3

Chang, D. H. F., & Troje, N. F. (2009a). Acceleration carries the local inversion effect in biological motion perception. *Journal of Vision, 9*(1).

Chang, D. H. F., & Troje, N. F. (2009b). Characterizing global and local mechanisms in biological motion perception. *Journal of Vision, 9*(5). doi:10.1167/9.5.8

Chartrand, T. L., & Bargh, J. A. (1999). The Chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76*(6), 893-910. doi:Doi 10.1037//0022-3514.76.6.893

Chong, S. C., Tadin, D., & Blake, R. (2005). Endogenous attention prolongs dominance durations in binocular rivalry. *J Vis, 5*(11), 1004-1012. doi:10.1167/5.11.6

Chouchourelou, A., Matsuka, T., Harber, K., & Shiffrar, M. (2006). The visual analysis of emotional actions. *Social Neuroscience, 1*(1), 63-74. doi:Doi 10.1080/17470910600630599

Ciaramidaro, A., Becchio, C., Colle, L., Bara, B. G., & Walter, H. (2014). Do you mean me? Communicative intentions recruit the mirror and the mentalizing system. *Soc Cogn Affect Neurosci, 9*(7), 909-916. doi:10.1093/scan/nst062

Clarke, T. J., Bradshaw, M. F., Field, D. T., Hampson, S. E., & Rose, D. (2005). The perception of emotion from body movement in point-light displays of interpersonal dialogue. *Perception, 34*(10), 1171-1180. doi:Doi 10.1068/P5203

Coelho, C. M., Cloete, S., & Wallis, G. (2010). The face-in-the-crowd effect: when angry faces are just cross(es). *J Vis, 10*(1), 7 1-14. doi:10.1167/10.1.7

Crammond, D. J., & Kalaska, J. F. (2000). Prior information in motor and premotor cortex: Activity during the delay period and effect on pre-movement activity. *Journal of Neurophysiology, 84*(2), 986-1005.

Crick, F., & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature, 375*(6527), 121-123. doi:10.1038/375121a0

Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *Journal of Cognitive Neuroscience, 17*(4), 641-651.

Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing Friends by Their Walk: Gait Perception without Familiarity Cues. *Bulletin of the Psychonomic Society, 9*(5), 353-356.

Cutting, J. E., Moore, C., & Morrison, R. (1988). Masking the Motions of Human Gait. *Perception & Psychophysics, 44*(4), 339-347.

Dayan, E., Casile, A., Levit-Binnun, N., Giese, M. A., Hendler, T., & Flash, T. (2007). Neural representations of kinematic laws of motion: evidence for action-perception coupling. *Proc Natl Acad Sci U S A, 104*(51), 20582-20587. doi:10.1073/pnas.0710033104

de Lange, F. P., Spronk, M., Willems, R. M., Toni, I., & Bekkering, H. (2008). Complementary systems for understanding action intentions. *Curr Biol, 18*(6), 454-457. doi:10.1016/j.cub.2008.02.057

Decety, J., & Grezes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Sciences, 3*(5), 172-178. doi:Doi 10.1016/S1364-6613(99)01312-1

Decety, J., & Jeannerod, M. (1995). Mentally simulated movements in virtual reality: does Fitts's law hold in motor imagery? *Behav Brain Res, 72*(1-2), 127-134.

DeLucia, P. R., & Liddell, G. W. (1998). Cognitive motion extrapolation and cognitive clocking in prediction motion tasks. *Journal of Experimental Psychology-Human Perception and Performance, 24*(3), 901-914. doi:Doi 10.1037/0096-1523.24.3.901

Devalois, R. L., & Devalois, K. K. (1991). Vernier Acuity with Stationary Moving Gabors. *Vision Research, 31*(9), 1619-1626.

Dipellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding Motor Events - a Neurophysiological Study. *Experimental Brain Research, 91*(1), 176-180.

Dittrich, W. H. (1993). Action categories and the perception of biological motion. *Perception, 22*(1), 15-22.

Dittrich, W. H., Troscianko, T., Lea, S. E. G., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception, 25*(6), 727-738. doi:10.1068/P250727

Dorris, M. C., Pare, M., & Munoz, D. P. (2000). Immediate neural plasticity shapes motor performance. *Journal of Neuroscience, 20*(1).

Dow, B. M., Snyder, A. Z., Vautin, R. G., & Bauer, R. (1981). Magnification Factor and Receptive-Field Size in Foveal Striate Cortex of the Monkey. *Experimental Brain Research, 44*(2), 213-228.

Downing, P. E., Jiang, Y. H., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science, 293*(5539), 2470-2473. doi:DOI 10.1126/science.1063414

Eagleman, D. M., & Sejnowski, T. J. (2000). Motion integration and postdiction in visual awareness. *Science, 287*(5460), 2036-2038.

Eifuku, S., & Wurtz, R. H. (1998). Response to motion in extrastriate area MSTl: center-surround interactions. *J Neurophysiol, 80*(1), 282-296.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience, 2*(10), 704-716. doi:Doi 10.1038/35094565

Eskenazi, T., Grosjean, M., Humphreys, G. W., & Knoblich, G. (2009). The role of motor simulation in action perception: a neuropsychological case study. *Psychological Research-Psychologische Forschung, 73*(4), 477-485. doi:DOI 10.1007/s00426-009-0231-5

Finke, R. A., Freyd, J., & Shyl, G. C. W. (1985). Memory Distortions Induced by Implied Velocity and Acceleration. *Bulletin of the Psychonomic Society, 23*(4), 276-276.

Finke, R. A., Freyd, J. J., & Shyi, G. C. W. (1986). Implied Velocity and Acceleration Induce Transformations of Visual Memory. *Journal of Experimental Psychology-General, 115*(2), 175-188. doi:Doi 10.1037//0096-3445.115.2.175

Flach, R., Knoblich, G., & Prinz, W. (2004). Recognizing one's own clapping: the role of temporal cues. *Psychol Res, 69*(1-2), 147-156. doi:10.1007/s00426-003-0165-2

Fox, R., & Mcdaniel, C. (1982). The Perception of Biological Motion by Human Infants. *Science, 218*(4571), 486-487.

Frantz, C. M., & Janoff-Bulman, R. (2000). Considering both sides: The limits of perspective taking. *Basic and Applied Social Psychology, 22*(1), 31-42. doi:Doi 10.1207/S15324834basp2201_4

Freyd, J. J. (1987). Dynamic Mental Representations. *Psychological Review, 94*(4), 427-438. doi:Doi 10.1037//0033-295x.94.4.427

Freyd, J. J., & Finke, R. A. (1984). Representational Momentum. *Journal of Experimental Psychology-Learning Memory and Cognition, 10*(1), 126-132.

Frith, C. D., & Frith, U. (1999). Interacting minds--a biological basis. *Science, 286*(5445), 1692-1695.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron, 50*(4), 531-534. doi:10.1016/j.neuron.2006.05.001

Fukushima, K. (1980). Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern, 36*(4), 193-202.

Gallant, J. L., Braun, J., & Vanessen, D. C. (1993). Selectivity for Polar, Hyperbolic, and Cartesian Gratings in Macaque Visual-Cortex. *Science, 259*(5091), 100-103. doi:DOI 10.1126/science.8418487

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*, 593-609. doi:DOI 10.1093/brain/119.2.593

Gawne, T. J., & Martin, J. M. (2002). Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. *J Neurophysiol, 88*(3), 1128-1135.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience, 4*(3), 179-192. doi:Doi 10.1038/Nrn1057

Gil-Da-Costa, R., Braun, A., Lopes, M., Hauser, M. D., Carson, R. E., Herscovitch, P., & Martin, A. (2004). Toward an evolutionary perspective on conceptual representation: Species-specific calls activate visual and affective processing systems in the macaque. *Proceedings of the National Academy of Sciences of the United States of America, 101*(50), 17516-17521. doi:DOI 10.1073/pnas.0408077101

Graf, M., Reitzner, B., Corves, C., Casile, A., Giese, M., & Prinz, W. (2007). Predicting point-light actions in real-time. *Neuroimage, 36*, T22-T32.

Grezes, J., Frith, C. D., & Passingham, R. E. (2004). Inferring false beliefs from the actions of oneself and others: an fMRI study. *Neuroimage, 21*(2), 744-750. doi:DOI 10.1016/j.neuroimage.2003.10.014

Grosjean, M., Shiffrar, M., & Knoblich, G. (2007). Fitts's law holds for action perception. *Psychol Sci, 18*(2), 95-99. doi:10.1111/j.1467-9280.2007.01854.x

Grossman, E. D., & Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron, 35*(6), 1167-1175. doi:Doi 10.1016/S0896-6273(02)00897-8

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences, 27*(3), 377-+.

Gunns, R. E., Johnston, L., & Hudson, S. M. (2002). Victim selection and kinematics: A point-light investigation of vulnerability to attack. *Journal of Nonverbal Behavior, 26*(3), 129-158.

Hegde, J., & Van Essen, D. C. (2000). Selectivity for complex shapes in primate visual area V2. *Journal of Neuroscience, 20*(5), art. no.-RC61.

Hikosaka, O., Sakamoto, M., & Usui, S. (1989). Functional-Properties of Monkey Caudate Neurons .3. Activities Related to Expectation of Target and Reward. *Journal of Neurophysiology, 61*(4), 814-832.

Hirai, M., Chang, D. H. F., Saunders, D. R., & Troje, N. F. (2011). Body Configuration Modulates the Usage of Local Cues to Direction in Biological-Motion Perception. *Psychological Science, 22*(12), 1543-1549. doi:Doi 10.1177/0956797611417257

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*(5), 1096-1106. doi:10.1037/0096-1523.31.5.1096

Hogeveen, J., & Obhi, S. S. (2012). Social Interaction Enhances Motor Resonance for Observed Human Actions. *Journal of Neuroscience, 32*(17), 5984-5989. doi:Doi 10.1523/Jneurosci.5938-11.2012

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol, 160*, 106-154.

Iacoboni, M., Lieberman, M. D., Knowlton, B. J., Molnar-Szakacs, I., Moritz, M., Throop, C. J., & Fiske, A. P. (2004). Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. *Neuroimage, 21*(3), 1167-1173. doi:10.1016/j.neuroimage.2003.11.013

Ichikawa, M., & Masakura, Y. (2006). Manual control of the visual stimulus reduces the flash-lag effect. *Vision Research, 46*(14), 2192-2203. doi:10.1016/j.visres.2005.12.021

Ichikawa, M., & Masakura, Y. (2010). Reduction of the flash-lag effect in terms of active observation. *Atten Percept Psychophys, 72*(4), 1032-1044. doi:10.3758/APP.72.4.1032

Jansson, G., & Johansso.G. (1973). Visual Perception of Bending Motion. *Perception, 2*(3), 321-326.

Jarraya, M., Amorim, M. A., & Bardy, B. G. (2005). Optical flow and viewpoint change modulate the perception and memorization of complex motion. *Perception & Psychophysics, 67*(6), 951-961. doi:Doi 10.3758/Bf03193622

Jellema, T., & Perrett, D. I. (2003). Cells in monkey STS responsive to articulated body motions and consequent static posture: a case of implied motion? *Neuropsychologia, 41*(13), 1728-1737. doi:Doi 10.1016/S0028-3932(03)00175-1

Jiang, Y., Costello, P., & He, S. (2007). Processing of invisible stimuli: advantage of upright faces and recognizable words in overcoming interocular suppression. *Psychol Sci, 18*(4), 349-355. doi:10.1111/j.1467-9280.2007.01902.x

Johansson, G. (1973). Visual-Perception of Biological Motion and a Model for Its Analysis. *Perception & Psychophysics, 14*(2), 201-211. doi:Doi 10.3758/Bf03212378

Jokisch, D., Daum, I., & Troje, N. F. (2006). Self recognition versus recognition of others by biological motion: Viewpoint-dependent effects. *Perception, 35*(7), 911-920. doi:Doi 10.1068/P5540

Jordan, H., Reiss, J. E., Hoffman, J. E., & Landau, B. (2002). Intact perception of biological motion in the face of profound spatial deficits: Williams syndrome. *Psychological Science, 13*(2), 162-167. doi:Doi 10.1111/1467-9280.00429

Kaiser, M. D., & Shiffrar, M. (2009). The visual perception of motion by observers with autism spectrum disorders: a review and synthesis. *Psychon Bull Rev, 16*(5), 761-777. doi:10.3758/PBR.16.5.761

Kanai, R., Sheth, B. R., & Shimojo, S. (2004). Stopping the motion and sleuthing the flash-lag effect: spatial uncertainty is the key to perceptual mislocalization. *Vision Research, 44*(22), 2605-2619.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron, 22*(4), 751-761. doi:Doi 10.1016/S0896-6273(00)80734-5

Kessler, K., Gordon, L., Cessford, K., & Lages, M. (2010). Characteristics of motor resonance predict the pattern of flash-lag effects for biological motion. *Plos One, 5*(1), e8258. doi:10.1371/journal.pone.0008258

Keysers, C., & Gazzola, V. (2007). Integrating simulation and theory of mind: from self to social cognition. *Trends Cogn Sci, 11*(5), 194-196. doi:10.1016/j.tics.2007.02.002

Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action: theory and data. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (pp. 59-101). Burlington: Academic Press.

Kohler, P. J., Cavanagh, P., & Tse, P. U. (2015). Motion-induced position shifts are influenced by global motion, but dominated by component motion. *Vision Research, 110*, 93-99.

Kourtis, D., Sebanz, N., & Knoblich, G. (2010). Favouritism in the motor system: social interaction modulates action simulation. *Biology Letters, 6*(6), 758-761. doi:DOI 10.1098/rsbl.2010.0478

Kourtzi, Z., & Kanwisher, N. (2000). Implied motion activates extrastriate motion-processing areas - Response to David and Senior (2000). *Trends in Cognitive Sciences, 4*(8), 295-296.

Kourtzi, Z., & Shiffrar, M. (1999). Dynamic representations of human body movement. *Perception, 28*(1), 49-62.

Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing Sex of a Walker from a Dynamic Point-Light Display. *Perception & Psychophysics, 21*(6), 575-580.

Krekelberg, B., & Lappe, M. (2000). A model of the perceived relative positions of moving objects based upon a slow averaging process. *Vision Research, 40*(2), 201-215.

Lagae, L., Raiguel, S., & Orban, G. A. (1993). Speed and Direction Selectivity of Macaque Middle Temporal Neurons. *Journal of Neurophysiology, 69*(1), 19-39.

Lampl, I., Ferster, D., Poggio, T., & Riesenhuber, M. (2004). Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex. *Journal of Neurophysiology, 92*(5), 2704-2713. doi:DOI 10.1152/jn.00060.2004

Lange, J., Georg, K., & Lappe, M. (2006). Visual perception of biological motion by form: a template-matching analysis. *J Vis, 6*(8), 836-849. doi:10.1167/6.8.6

Lange, J., & Lappe, M. (2006). A model of biological motion perception from configural form cues. *J Neurosci, 26*(11), 2894-2906. doi:10.1523/JNEUROSCI.4915-05.2006

Lappe, M., & Krekelberg, B. (1998). The position of moving objects. *Perception, 27*(12), 1437-1449.

Levelt, W. J. M. (1968). *On binocular rivalry*. Hague,: Mouton.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape Representation in the Inferior Temporal Cortex of Monkeys. *Current Biology, 5*(5), 552-563. doi:Doi 10.1016/S0960-9822(95)00108-4

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience, 19*, 577-621. doi:DOI 10.1146/annurev.neuro.19.1.577

Loula, F., Prasad, S., Harber, K., & Shiffrar, M. (2005). Recognizing people from their movement. *Journal of Experimental Psychology-Human Perception and Performance, 31*(1), 210-220. doi:Doi 10.1037/0096-1523.31.1.210

Lu, H. (2010). Structural processing in biological motion perception. *Journal of Vision, 10*(12). doi:Artn 13

Doi 10.1167/10.12.13

Lu, H., Tjan, B. S., & Liu, Z. (2006). Shape recognition alters sensitivity in stereoscopic depth discrimination. *J Vis, 6*(1), 75-86. doi:10.1167/6.1.7

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology, 77*(1), 24-42.

Lyon, D. R., & Waag, W. L. (1995). Time-Course of Visual Extrapolation Accuracy. *Acta Psychologica, 89*(3), 239-260. doi:Doi 10.1016/0001-6918(95)98945-Z

Mackay, D. M. (1958). Perceptual Stability of a Stroboscopically Lit Visual Field Containing Self-Luminous Objects. *Nature, 181*(4607), 507-508.

Malle, B. F., & Pearce, G. E. (2001). Attention to behavioral events during interaction: Two actor-observer gaps and three attempts to close them. *Journal of Personality and Social Psychology, 81*(2), 278-294. doi:10.1037//0022-3514.81.2.278

Manera, V., Becchio, C., Schouten, B., Bara, B. G., & Verfaillie, K. (2011). Communicative Interactions Improve Visual Detection of Biological Motion. *Plos One, 6*(1). doi:10.1371/journal.pone.0014594

Manera, V., Schouten, B., Verfaillie, K., & Becchio, C. (2013). Time Will Show: Real Time Predictions during Interpersonal Action Perception. *Plos One, 8*(1).

Marsh, A. A., Kozak, M. N., Wegner, D. M., Reid, M. E., Yu, H. H., & Blair, R. J. R. (2010). The neural substrates of action identification. *Social Cognitive and Affective Neuroscience, 5*(4), 392-403. doi:10.1093/scan/nsq004

Mather, G., & Murdoch, L. (1994). Gender Discrimination in Biological Motion Displays Based on Dynamic Cues. *Proceedings of the Royal Society B: Biological Sciences, 258*(1353), 273-279. doi:10.1098/rspb.1994.0173

Maus, G. W., & Nijhawan, R. (2008). Motion extrapolation into the blind spot. *Psychological Science, 19*(11), 1087-1091.

Maus, G. W., & Nijhawan, R. (2009). Going, going, gone: localizing abrupt offsets of moving objects. *J Exp Psychol Hum Percept Perform, 35*(3), 611-626. doi:10.1037/a0012317

Maus, G. W., Ward, J., Nijhawan, R., & Whitney, D. (2013). The perceived position of moving objects: transcranial magnetic stimulation of area MT+ reduces the flash-lag effect. *Cereb Cortex, 23*(1), 241-247. doi:10.1093/cercor/bhs021

McLeod, P., Dittrich, W., Driver, J., Perrett, D., & Zihl, J. (1996). Preserved and impaired detection of structure from motion by a "motion-blind" patient. *Visual Cognition, 3*(4), 363-391. doi:Doi 10.1080/135062896395634

Mel, B. W., & Fiser, J. (2000). Minimizing binding errors using learned conjunctive features. *Neural Computation, 12*(2), 247-278. doi:Doi 10.1162/089976600300015772

Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews Neuroscience, 1*(1), 59-65. doi:Doi 10.1038/35036228

Miller, L. E., & Saygin, A. P. (2013). Individual differences in the perception of biological motion: links to social cognition and motor imagery. *Cognition, 128*(2), 140-148. doi:10.1016/j.cognition.2013.03.013

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action* (1st ed.). Oxford ; New York: Oxford University Press.

Montepare, J. M., & Zebrowitzmcarthur, L. (1988). Impressions of People Created by Age-Related Qualities of Their Gaits. *Journal of Personality and Social Psychology, 55*(4), 547-556.

Morillon, B., Lehongre, K., Frackowiak, R. S. J., Ducorps, A., Kleinschmidt, A., Poeppel, D., & Giraud, A. L. (2010). Neurophysiological origin of human brain asymmetry for speech and language. *Proceedings of the National Academy of Sciences of the United States of America, 107*(43), 18688-18693. doi:DOI 10.1073/pnas.1007189107

Neri, P., Luu, J. Y., & Levi, D. M. (2006). Meaningful interactions can enhance visual discrimination of human agents. *Nature Neuroscience, 9*(9), 1186-1192. doi:10.1038/Nn1759

Neri, P., Morrone, M. C., & Burr, D. C. (1998). Seeing biological motion. *Nature, 395*(6705), 894-896.

Nijhawan, R. (1994). Motion Extrapolation in Catching. *Nature, 370*(6487), 256-257.

Nijhawan, R. (2001). The flash-lag phenomenon: object motion and eye movements. *Perception, 30*(3), 263-282.

Nijhawan, R. (2008). Visual prediction: Psychophysics and neurophysiology of compensation for time delays. *Behavioral and Brain Sciences, 31*(02), 179-198.

Nijhawan, R., & Kirschfeld, K. (2003). Analogous mechanisms compensate for neural delays in the sensory and the motor pathways: evidence from motor flash-lag. *Curr Biol, 13*(9), 749-753.

Obhi, S. S., & Sebanz, N. (2011). Moving together: toward understanding the mechanisms of joint action. *Exp Brain Res, 211*(3-4), 329-336. doi:10.1007/s00221-011-2721-0

Olson, I. R., Gatenby, J. C., Leung, H. C., Skudlarski, P., & Gore, J. C. (2004). Neuronal representation of occluded objects in the human brain. *Neuropsychologia, 42*(1), 95-104. doi:Doi 10.1016/S0028-3932(03)00151-9

Orban, G. A., Dupont, P., De Bruyn, B., Vogels, R., Vandenberghe, R., & Mortelmans, L. (1995). A motion area in human visual cortex. *Proc Natl Acad Sci U S A, 92*(4), 993-997.

Patel, S. S., Ogmen, H., Bedell, H. E., & Sampath, V. (2000). Flash-lag effect: differential latency, not postdiction. *Science, 290*(5494), 1051.

Pavlova, M., & Sokolov, A. (2000). Orientation specificity in biological motion perception. *Perception & Psychophysics, 62*(5), 889-899.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis, 10*(4), 437-442.

Perrett, D. I., & Oram, M. W. (1993). Neurophysiology of Shape Processing. *Image and Vision Computing, 11*(6), 317-333. doi:Doi 10.1016/0262-8856(93)90011-5

Perrett, D. I., Smith, P. A. J., Mistlin, A. J., Chitty, A. J., Head, A. S., Potter, D. D., . . . Jeeves, M. A. (1985). Visual Analysis of Body Movements by Neurons in the Temporal Cortex of the Macaque Monkey - a Preliminary-Report. *Behavioural Brain Research, 16*(2-3), 153-170. doi:Doi 10.1016/0166-4328(85)90089-0

Pinto, J., & Shiffrar, M. (1999). Subconfigurations of the human form in the perception of biological motion displays. *Acta Psychologica, 102*(2-3), 293-318.

Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition, 82*(2), B51-B61.

Prinz, W. (2006). What re-enactment earns us. *Cortex, 42*(4), 515-517.

Proffitt, D. R., & Bertenthal, B. I. (1990). Converging Operations Revisited - Assessing What Infants Perceive Using Discrimination Measures. *Perception & Psychophysics, 47*(1), 1-11. doi:Doi 10.3758/Bf03208159

Purushothaman, G., Patel, S. S., Bedell, H. E., & Ogmen, H. (1998). Moving ahead through differential visual latency. *Nature, 396*(6710), 424-424.

Rao, R. P., Eagleman, D. M., & Sejnowski, T. J. (2001). Optimal smoothing in visual motion perception. *Neural Comput, 13*(6), 1243-1253.

Reed, C. L., Stone, V. E., Bozova, S., & Tanaka, J. (2003). The body-inversion effect. *Psychological Science, 14*(4), 302-308. doi:Doi 10.1111/1467-9280.14431

Rees, D., Backus, B. T., & Heeger, D. J. (2000). Activity in primary visual cortex predicts performance in a visual detection task (vol 3, pg 940, 2000). *Nature Neuroscience, 3*(12), 1347-1347.

Rice, A., Phillips, P. J., Natu, V., An, X., & O'Toole, A. J. (2013). Unaware person recognition

    from the body when face identification fails. *Psychol Sci, 24*(11), 2235-2243.

    doi:10.1177/0956797613492986

Riehle, A., & Requin, J. (1993). The Predictive Value for Performance Speed of Preparatory

    Changes in Neuronal-Activity of the Monkey Motor and Premotor Cortex. *Behavioural*

    Brain Research, 53(1-2), 35-49. doi:Doi 10.1016/S0166-4328(05)80264-5

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex.

    Nature Neuroscience, 2(11), 1019-1025.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the

    understanding and imitation of action. *Nature Reviews Neuroscience, 2*(9), 661-670.

    doi:Doi 10.1038/35090060

Rodman, H. R., & Albright, T. D. (1989). Single-Unit Analysis of Pattern-Motion Selective

    Properties in the Middle Temporal Visual Area (Mt). *Experimental Brain Research,*

    75(1), 53-64.

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). doi:10.1167/9.6.15

Rosenbaum, D. A. (1975). Perception and Extrapolation of Velocity and Acceleration. *Journal of*

    Experimental Psychology-Human Perception and Performance, 1(4), 395-403. doi:Doi

    10.1037/0096-1523.1.4.395

Roulston, B. W., Self, M. W., & Zeki, S. (2006). Perceptual compression of space through

    position integration. *Proc Biol Sci, 273*(1600), 2507-2512. doi:10.1098/rspb.2006.3616

Runeson, S., & Frykholm, G. (1983). Kinematic Specification of Dynamics as an Informational Basis for Person-and-Action Perception - Expectation, Gender Recognition, and Deceptive Intention. *Journal of Experimental Psychology-General, 112*(4), 585-615.

Rutherford, M. D., & Kuhlmeier, V. A. (2013). *Social perception detection and interpretation of animacy, agency, and intention* (pp. 1 online resource (vi, 415 p.)). Retrieved from http://dx.doi.org/10.7551/mitpress/9780262019279.001.0001

Saxe, R., Xiao, D. K., Kovacs, G., Perrett, D. I., & Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia, 42*(11), 1435-1446. doi:10.1016/j.neuropsychologia.2004.04.015

Schiff, W., & Detwiler, M. L. (1979). Information Used in Judging Impending Collision. *Perception, 8*(6), 647-658. doi:Doi 10.1068/P080647

Schiff, W., & Oldak, R. (1990). Accuracy of Judging Time to Arrival - Effects of Modality, Trajectory, and Gender. *Journal of Experimental Psychology-Human Perception and Performance, 16*(2), 303-316. doi:Doi 10.1037/0096-1523.16.2.303

Scocchia, L., Grosso, R. A., de'Sperati, C., Stucchi, N., & Baud-Bovy, G. (2009). Observer's control of the moving stimulus increases the flash-lag effect. *Vision Research, 49*(19), 2363-2370. doi:10.1016/j.visres.2009.06.023

Senior, C., Barnes, J., Giampietro, V., Simmons, A., Bullmore, E. T., Brammer, M., & David, A. S. (2000). The functional neuroanatomy of implicit-motion perception or 'representational momentum'. *Current Biology, 10*(1), 16-22.

Senior, C., Ward, J., & David, A. S. (2002). 'Representational momentum and the brain: An investigation into the functional necessity of V5/MT' (vol 9, pg 81, 2002). *Visual Cognition, 9*(3), 382-382. doi:Doi 10.1080/13506280244000032

Sevdalis, V., & Keller, P. E. (2012). Perceiving bodies in motion: expression intensity, empathy, and experience. *Experimental Brain Research, 222*(4), 447-453. doi:DOI 10.1007/s00221-012-3229-y

Shiffrar, M., Lichtey, L., & Chatterjee, S. H. (1997). The perception of biological motion across apertures. *Perception & Psychophysics, 59*(1), 51-59.

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences of the United States of America, 105*(2), 809-813. doi:DOI 10.1073/pnas.0707021105

Sobel, K. V., & Blake, R. (2002). How context influences predominance during binocular rivalry. *Perception, 31*(7), 813-824.

Sparenberg, P., Springer, A., & Prinz, W. (2012). Predicting others' actions: evidence for a constant time delay in action simulation. *Psychological Research-Psychologische Forschung, 76*(1), 41-49. doi:10.1007/s00426-011-0321-z

Spence, C., & Driver, J. (2000). Attracting attention to the illusory location of a sound: reflexive crossmodal orienting and ventriloquism. *Neuroreport, 11*(9), 2057-2061.

Sperduti, M., Guionnet, S., Fossati, P., & Nadel, J. (2014). Mirror Neuron System and Mentalizing System connect during online social interaction. *Cogn Process, 15*(3), 307-316. doi:10.1007/s10339-014-0600-x

Springer, A., & Prinz, W. (2010). Action semantics modulate action prediction. *Quarterly Journal of Experimental Psychology, 63*(11), 2141-2158. doi:10.1080/17470211003721659

Spunt, R. P., & Lieberman, M. D. (2013). The busy social brain: evidence for automaticity and

    control in the neural systems supporting social cognition and action understanding.

    Psychol Sci, 24(1), 80-86. doi:10.1177/0956797612450884

Stadler, W., Springer, A., Parkinson, J., & Prinz, W. (2012). Movement kinematics affect action

    prediction: comparing human to non-human point-light actions. *Psychological Research-*

    Psychologische Forschung, 76(4), 395-406. doi:DOI 10.1007/s00426-012-0431-2

Stein, T., Senju, A., Peelen, M. V., & Sterzer, P. (2011). Eye contact facilitates awareness of

    faces during interocular suppression. *Cognition, 119*(2), 307-311.

    doi:10.1016/j.cognition.2011.01.008

Sumi, S. (1984). Upside-down presentation of the Johansson moving light-spot pattern.

    Perception, 13(3), 283-286.

Sumi, S. (2000). Perception of point-light walker produced by eight lights attached to the back of

    the walker. *Swiss Journal of Psychology, 59*(2), 126-132.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in*

    Cognitive Sciences, 13(9), 403-409. doi:DOI 10.1016/j.tics.2009.06.003

Tanaka, K., & Saito, H. A. (1989). Analysis of Motion of the Visual-Field by Direction,

    Expansion Contraction, and Rotation Cells Clustered in the Dorsal Part of the Medial

    Superior Temporal Area of the Macaque Monkey. *Journal of Neurophysiology, 62*(3),

    626-641.

Thompson, B., Hansen, B. C., Hess, R. F., & Troje, N. F. (2007). Peripheral vision: Good for

    biological motion, bad for signal noise segregation? *Journal of Vision, 7*(10).

    doi:10.1167/7.10.12

Thornton, I. M., & Hayes, A. E. (2004). Anticipating action in complex scenes. *Visual Cognition, 11*(2-3), 341-370. doi:10.1080/13506280344000374

Thurman, S. M., & Lu, H. (2013). Physical and biological constraints govern perceived animacy of scrambled human forms. *Psychol Sci, 24*(7), 1133-1141. doi:10.1177/0956797612467212

Thurman, S. M., & Lu, H. (2014). Perception of social interactions for spatially scrambled biological motion. *Plos One, 9*(11), e112539. doi:10.1371/journal.pone.0112539

Troje, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision, 2*(5), 371-387. doi:Doi 10.1167/2.5.2

Troje, N. F., & Geyer, H. (2002). Aftereffects in biological motion perception. *Perception, 31*, 152-152.

Troje, N. F., & Westhoff, C. (2006). The inversion effect in biological motion perception: Evidence for a "life detector"? *Current Biology, 16*(8), 821-824. doi:DOI 10.1016/j.cub.2006.03.022

Troje, N. F., Westhoff, C., & Lavrov, M. (2005). Person identification from biological motion: Effects of structural and kinematic cues. *Perception & Psychophysics, 67*(4), 667-675. doi:Doi 10.3758/Bf03193523

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nat Neurosci, 8*(8), 1096-1101. doi:10.1038/nn1500

Vaina, L. M., Lemay, M., Bienfang, D. C., Choi, A. Y., & Nakayama, K. (1990). Intact Biological Motion and Structure from Motion Perception in a Patient with Impaired Motion Mechanisms - a Case-Study. *Visual Neuroscience, 5*(4), 353-369.

van Boxtel, J. J., Alais, D., & van Ee, R. (2008). Retinotopic and non-retinotopic stimulus encoding in binocular rivalry and the involvement of feedback. *J Vis, 8*(5), 17 11-10. doi:10.1167/8.5.17

van Boxtel, J. J., & Lu, H. (2012). Signature movements lead to efficient search for threatening actions. *Plos One, 7*(5), e37085. doi:10.1371/journal.pone.0037085

van Boxtel, J. J., & Lu, H. (2013a). A biological motion toolbox for reading, displaying, and manipulating motion capture data in research settings. *Journal of Vision, 13*(12).

van Boxtel, J. J., & Lu, H. (2013b). Impaired global, and compensatory local, biological motion processing in people with high levels of autistic traits. *Front Psychol, 4*, 209. doi:10.3389/fpsyg.2013.00209

van Boxtel, J. J., & Lu, H. (2015a). Joints and their relations as critical features in action discrimination: evidence from a classification image method. *J Vis, 15*(1), 15 11 20. doi:10.1167/15.1.20

van Boxtel, J. J., & Lu, H. (2015b). Understanding Biological Motion. In R. A. Scott & S. M. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource* (pp. 1-14). Hoboken, NJ: John Wiley & Sons, Inc.,.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage, 48*(3), 564-584. doi:10.1016/j.neuroimage.2009.06.009

Verfaillie, K., & Daems, A. (2002). Representing and anticipating human actions in vision. *Visual Cognition, 9*(1-2), 217-232. doi:Doi 10.1080/13506280143000403

Walk, R. D., & Homan, C. P. (1984). Emotion and Dance in Dynamic Light Displays. *Bulletin of the Psychonomic Society, 22*(5), 437-440.

Wang, L., & Jiang, Y. (2012). Life motion signals lengthen perceived temporal duration. *Proc Natl Acad Sci U S A, 109*(11), E673-677. doi:10.1073/pnas.1115515109

Watson, T. L., Pearson, J., & Clifford, C. W. (2004). Perceptual grouping of biological motion promotes binocular rivalry. *Curr Biol, 14*(18), 1670-1674. doi:10.1016/j.cub.2004.08.064

Wheaton, K. J., Thompson, J. C., Syngeniotis, A., Abbott, D. F., & Puce, A. (2004). Viewing the motion of human body parts activates different regions of premotor, temporal, and parietal cortex. *Neuroimage, 22*(1), 277-288. doi:DOI 10.1016/j.neuroimage.2003.12.043

Whitney, D. (2002). The influence of visual motion on perceived position. *Trends in Cognitive Sciences, 6*(5), 211-216. doi:Doi 10.1016/S1364-6613(02)01887-9

Whitney, D., & Cavanagh, P. (2000). Motion distorts visual space: shifting the perceived position of remote stationary objects. *Nature Neuroscience, 3*(9), 954-959.

Whitney, D., Murakami, I., & Cavanagh, P. (2000). Illusory spatial offset of a flash relative to a moving stimulus is caused by differential latencies for moving and flashed stimuli. *Vision Research, 40*(2), 137-149.

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin, 131*(3), 460-473. doi:10.1037/0033-2909.131.3.460

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society B-Biological Sciences, 358*(1431), 593-602. doi:DOI 10.1098/rstb.2002.1238

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology, 11*(18), R729-R732.

Xiao, D. K., Raiguel, S., Marcar, V., Koenderink, J., & Orban, G. A. (1995). Spatial Heterogeneity of Inhibitory Surrounds in the Middle Temporal Visual Area. *Proceedings of the National Academy of Sciences of the United States of America, 92*(24), 11303-11306. doi:DOI 10.1073/pnas.92.24.11303

Yang, E., Zald, D. H., & Blake, R. (2007). Fearful expressions gain preferential access to awareness during continuous flash suppression. *Emotion, 7*(4), 882-886. doi:10.1037/1528-3542.7.4.882

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., . . . Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience, 4*(6), 651-655. doi:Doi 10.1038/88486