

Cross-modal social attention triggered by biological motion cues

Yiwen Yu

State Key Laboratory of Brain and Cognitive Science,
CAS Center for Excellence in Brain Science and
Intelligence Technology, Institute of Psychology, Chinese
Academy of Sciences, Beijing, China
Department of Psychology, University of Chinese Academy
of Sciences, Beijing, China
Chinese Institute for Brain Research, Beijing, China



Haoyue Ji

State Key Laboratory of Brain and Cognitive Science,
CAS Center for Excellence in Brain Science and
Intelligence Technology, Institute of Psychology, Chinese
Academy of Sciences, Beijing, China
Department of Psychology, University of Chinese Academy
of Sciences, Beijing, China
Chinese Institute for Brain Research, Beijing, China



Li Wang

State Key Laboratory of Brain and Cognitive Science,
CAS Center for Excellence in Brain Science and
Intelligence Technology, Institute of Psychology, Chinese
Academy of Sciences, Beijing, China
Department of Psychology, University of Chinese Academy
of Sciences, Beijing, China
Chinese Institute for Brain Research, Beijing, China



Yi Jiang

State Key Laboratory of Brain and Cognitive Science,
CAS Center for Excellence in Brain Science and
Intelligence Technology, Institute of Psychology, Chinese
Academy of Sciences, Beijing, China
Department of Psychology, University of Chinese Academy
of Sciences, Beijing, China
Chinese Institute for Brain Research, Beijing, China
Institute of Artificial Intelligence, Hefei Comprehensive
National Science Center, Hefei, China



Previous research has demonstrated that biological motion (BM) cues can induce reflexive attentional orienting. This BM-triggered social attention has hitherto only been investigated within visual modality. It remains unknown whether and to what extent social attention induced by BM cues can occur across different sensory modalities. By introducing auditory stimuli to a modified central cueing paradigm, we showed that observers responded significantly faster to auditory targets presented in the walking direction of BM than in the opposite direction, reflecting the notion that BM

cues can trigger cross-modal social attention. This effect was not due to the viewpoint effect of the global configuration and could be extended to local BM cues without any global configuration. Critically, such cross-modal social attention was sensitive to the orientation of BM cues and completely disappeared when critical biological characteristics were removed. Our findings, taken together, support the existence of a special multimodal attention mechanism tuned to life motion signals and shed new light on the unique and cross-modal nature of social attention.

Citation: Yu, Y., Ji, H., Wang, L., & Jiang, Y. (2020). Cross-modal social attention triggered by biological motion cues. *Journal of Vision*, 20(10):21, 1–10, <https://doi.org/10.1167/jov.20.10.21>.



Introduction

Humans appear to be endowed with the ability to readily compute the whereabouts of others' focus of interest via social cues (e.g., eye gaze) and use this information accordingly to change their own attentional behaviors (Birmingham & Kingstone, 2009; Frischen, Bayliss, & Tipper, 2007; Nummenmaa & Calder, 2009). This fundamental social attention ability is of prime importance for social interactions and adaptive functioning, as it enables humans to learn about other people's mental states (e.g., intentions, goals) and where the object of interest (e.g., food) is in the surrounding environment (Klein, Shepherd, & Platt, 2009). Such ability is an important precursor to language acquisition and theory-of-mind development (Baron-Cohen, 1995; R. Brooks & Meltzoff, 2005; Nuku & Bekkering, 2008; Shepherd, 2010). A modified central cueing task that is simple and well established has been widely used for simulating and measuring social attention (Friesen & Kingstone, 1998; Frischen, Bayliss, et al., 2007; Mathews, Fox, Yiend, & Calder, 2003). Typically, a nonpredictive eye gaze cue is presented centrally, and it will trigger a reflexive attentional orienting as evidenced by more rapid responses to targets appearing in a location congruent with the direction of the eye gaze than those in the opposite location (i.e., gaze cueing effect). This effect arises very rapidly (as early as 100 ms after the appearance of the gaze stimulus) and occurs even when eye gaze direction is counter-predictive of target location, thus disclosing its reflexive nature (Driver, Davis, Ricciardelli, Kidd, Maxwell, & Baron-Cohen, 1999; Friesen, Ristic, & Kingstone, 2004; Tipples, 2008). In this respect, social attention resembles the well-known exogenous attention that uses spatially uninformative cues to investigate the reflexive response of attention (Posner, 1980). However, unlike exogenous attention, social attention is not caused by the peripherally presented cue, and its inhibition of return effect is quite delayed (Frischen, Smilek, Eastwood, & Tipper, 2007). On the other hand, although social cues appear in the central location like traditional endogenous cues (Posner, 1980), the direction of social cues is not predictive of the probable location of the subsequent target, thus distinguishing social attention from endogenous attention. Given these special properties, this type of reflexive attentional orienting challenges the traditional dichotomous categorization of covert attention and opens up new avenues for visual attention research.

In everyday life, eyes are not the unique source of information regarding others' focus of attention. The movements of biological organisms can also be used to identify the source of interest of a peer when the eyes are not visible (J. Thompson & Parasuraman, 2012).

Biological motion (BM), even depicted solely by several moving point-light dots attached to the major joints, suffices to provide quite rich meaningful information of biological and social relevance, such as gender (A. Brooks, Schouten, Troje, Verfaillie, Blanke, & van der Zwan, 2008), emotion (K. L. Johnson, McKay, & Pollick, 2011), intention (Manera, Schouten, Becchio, Bara, & Verfaillie, 2010), and so forth. Among them, walking direction is an especially important attribute of BM, as it conveys the disposition and goals of a biological entity. BM perception is therefore commonly considered to be a hallmark of social cognition (Pavlova, 2012) and appears to be compromised in individuals with social cognitive deficits (e.g., autism) (Klin, Lin, Gorrindo, Ramsay, & Jones, 2009).

Humans and a wide range of non-human species (e.g., chicks, monkeys) have evolved to be highly sensitive to the direction of BM (Di Giorgio, Loveland, Mayer, Rosa-Salva, Versace, & Vallortigara, 2017; Oram & Perrett, 1994; Rosa Salva, Mayer, & Vallortigara, 2015; Simion, Regolin, & Bulf, 2008; Sweeny, Wurnitsch, Gopnik, & Whitney, 2013; B. Thompson, Hansen, Hess, & Troje, 2007). For example, observers are able to discriminate the direction of a point-light walker even when it is masked by moving random dots (Bertenthal & Pinto, 1994). Moreover, this ability emerges very early in life (6-month-old infant) (Kuhlmeier, Troje, & Lee, 2010). Importantly, such intrinsic sensitivity to the direction of BM can further affect our attentional behaviors. A previous study has demonstrated that the walking direction of BM cues can trigger a reflexive orienting effect (Shi, Weng, He, & Jiang, 2010). This attentional effect can also be observed in preschool children and 6-month-old infants (Bardi, Di Giorgio, Lunghi, Troje, & Simion, 2015; Zhao, Wang, Wang, Weng, Li, & Jiang, 2014). Such BM-induced attentional orienting effect persists even when the global configuration of BM is disrupted and observers are naïve to its biological nature (L. Wang, Yang, Shi, & Jiang, 2014). Our recent study has further demonstrated that this social attention ability is highly heritable and hardwired in the human visual system (L. Wang et al., 2020), which is also in line with the findings from other species (Di Giorgio et al., 2017; Rosa Salva et al., 2015; Vallortigara & Regolin, 2006; Vallortigara, Regolin, & Marconato, 2005). Moreover, the attentional orienting effects triggered by BM and eye gaze cues share common genetic bases and neural mechanisms (Ji, Wang, & Jiang, 2020; L. Wang et al., 2020), which implies the existence of a “social attention detector” in the human brain.

Previous investigations into BM-triggered attention have specifically focused on one sense modality (i.e., vision). In other words, all of these studies employed the unimodal cueing task, in which visually presented BM cues facilitated the detection of visual targets. In the real world, however, the environment is much

more complex, and we constantly have to cope with sensory inputs belonging to different modalities. From an evolutionary perspective, it is essential to detect the occurrence of important events (e.g., food, danger) through non-visual modalities (e.g., audition), especially when such signals are out of sight or obscured. Therefore, it would be of great interest to investigate social attention induced by BM cues under cross-modal conditions. Until recently, the multisensory modulatory effects, particularly the modulation of auditory cues on visual BM, have become a focus on research into BM perception. It has been demonstrated that auditory cues can affect visual BM processing (e.g., direction, gender) (Arrighi, Marini, & Burr, 2009; A. Brooks, van der Zwan, Billard, Petreska, Clarke, & Blanke, 2007; Mendonca, Santos, & Lopez-Moliner, 2011; Schouten, Troje, Vroomen, & Verfaillie, 2011; Thomas & Shiffrar, 2010; van der Zwan, MacHatch, Kozlowski, Troje, Blanke, & Brooks, 2009; Wuerger, Crocker-Buque, & Meyer, 2012). For example, direction-matched auditory motion enhanced the direction discrimination performance of visually presented BM, whereas auditory motion in the opposite direction impeded the detection (A. Brooks et al., 2007). Moreover, auditory cues could also bias the gender perception of ambiguous BM. Specifically, female auditory cues made gender-ambiguous BM look more feminine (van der Zwan et al., 2009). However, it remains unclear whether visual BM cues can modulate attention to auditory information.

To probe this issue, the present study investigated whether the walking direction of BM could modulate auditory attention by using auditory targets in a modified central cueing paradigm. Similar to the previous cross-modal gaze cueing study (Newport & Howarth, 2009), we manipulated the BM stimulus as either walking toward left or right as a central cue, followed by a lateralized tone that was presented from headphones and served as the target. Participants were required to perform a sound localization task immediately after viewing the central BM cue. In addition to intact BM cues, the current study also employed feet motion sequences as central cues to investigate whether such cross-modal BM-triggered attention, if observed, could extend to local BM cues without global configuration.

Methods

Participants

Ninety-six participants (54 females) who between the ages of 18 and 35 years ($M \pm SD = 24.3 \pm 4.0$ years) were recruited in four experiments. Twenty-four (13 females) participated in Experiment 1, 24 (14 females)

in Experiment 2, 24 (14 females) in Experiment 3, and the remaining 24 (13 females) in Experiment 4. All participants had normal or corrected-to-normal vision and normal hearing and gave written informed consent in accordance with procedures and protocols approved by the institutional review board of the Institute of Psychology, Chinese Academy of Sciences. They were all naïve to the purpose of the experiments.

Stimuli and procedure

Stimuli were generated and displayed using MATLAB (MathWorks, Natick, MA) together with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997) on a 19-inch cathode-ray tube monitor (1280×1024 at 60 Hz). BM stimuli with leftward or rightward walking direction, depicted by 13 white point-light dots that represented the motions of the head and the major joints (shoulders, elbows, wrists, hips, knees, and ankles) of a walker, were adopted from Vanrie and Verfaillie (2004). Each cycle was 1 second and contained 30 frames. In each trial, to avoid observers' prediction, the initial frame of the point-light display was randomized. Static BM frames were created by capturing the most extended points of a gait cycle from the BM stimuli. Nonbiological motion sequences were derived from the fragments identical to the BM stimuli but with critical biological characteristics removed. Specifically, each individual dot moved along a path identical to that of the BM stimuli but with a constant speed equal to the average speed of all the dots. In addition, the initial motion phase of each individual dot was also randomized. Such manipulations disrupted the natural velocity profile and phase relationship of the BM stimuli but kept the motion trajectories of individual point lights unchanged. The feet motion sequences that served as local BM cues were created by isolating the two point lights of ankles from the original BM sequences. They consisted of two fragments representing the stance phase and swing phase of the foot trajectory. During the stance phase, the corresponding dot moved in the opposite direction of the walking direction at an approximately constant velocity. During the swing phase, the dot accelerated along both the horizontal and vertical dimensions due to muscle activity and gravitational acceleration. Note that the intact BM cues contained both global configuration information (i.e., skeletal structure) and local motion information (i.e., trajectories of individual dots), whereas the local BM cues were devoid of configural cues and retained local motion signals only. Inverted counterparts were created by mirror flipping the motion sequences (i.e., BM and feet motion sequences) vertically such that the walking

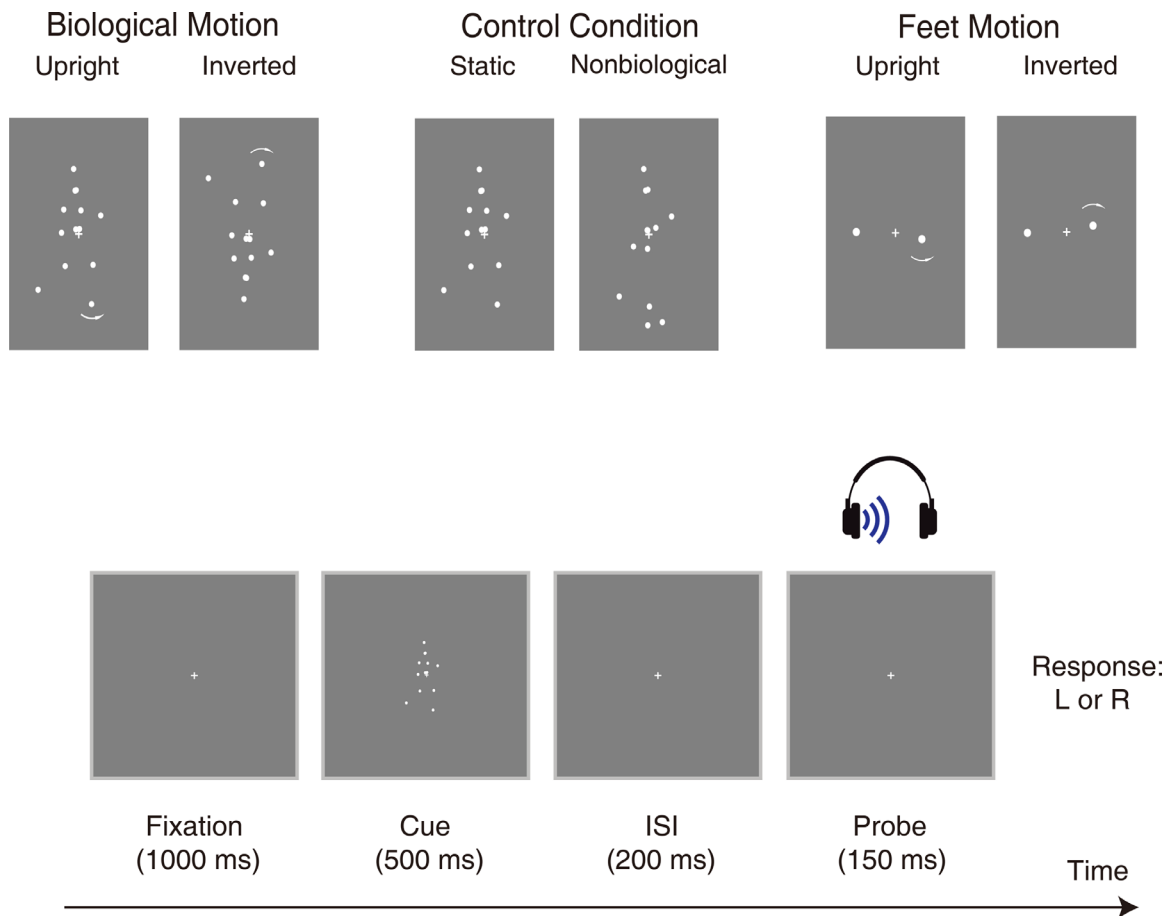


Figure 1. Static frames of sample stimuli used in Experiments 1 to 4 and schematic representation of the experimental paradigm. In Experiment 1, a point-light walker (upright or inverted) was presented as a central cue for 500 ms in each trial. After a 200-ms ISI, participants were required to indicate the location (left or right) of a briefly presented (150 ms) auditory probe. Experiments 2 to 4 followed a similar procedure, with the exception that the central cues were static BM frames in Experiment 2, nonbiological motion sequences in Experiment 3, and feet motion sequences (upright and inverted) in Experiment 4. At the beginning of each experiment, participants were explicitly told that the cue direction did not predict the probe location.

direction was kept the same for the upright and inverted versions.

In Experiment 1, all stimuli were presented within a white frame ($17.5^\circ \times 17.5^\circ$) on a gray background (RGB: 128, 128, 128), and the viewing distance was about 57 cm. Each trial began with a white cross ($0.6^\circ \times 0.6^\circ$) displayed in the center of the screen on which participants were asked to fixate throughout the experiment. After 1000 ms, an upright or inverted BM sequence (subtended approximately $3.5^\circ \times 6.0^\circ$ in visual angle) appeared as a central cue for 500 ms. After a 200-ms interstimulus interval (ISI) in which only the fixation was displayed, a pure tone was presented briefly (150 ms) as a probe to the left or right ear through the headphone. Participants were required to indicate the location (left or right) of the auditory probe as quickly and accurately as possible by pressing the left or right arrow key, respectively, on the keyboard (Figure 1). At the beginning of the experiment, participants were

explicitly told that the cue direction did not predict the probe location. There were two blocks, the upright BM stimulus block and the inverted BM stimulus block. Each block consisted of 80 trials with 40 congruent trials (the probe location was the same as the cued direction) and 40 incongruent trials (the probe location was opposite to the cued direction). Test trials were presented in a new random order for each participant. The order of the blocks (upright and inverted) was also counterbalanced across participants. Experiments 2 and 3 followed a procedure similar to that in the upright BM stimulus block of Experiment 1, except that static BM frames and nonbiological motion sequences were employed as central cues. Experiment 4 was identical to Experiment 1, with the exception that feet motion sequences (upright and inverted) were used as cues. In addition, participants were informed of the biological nature of upright and inverted feet motion sequences before the experiment began.

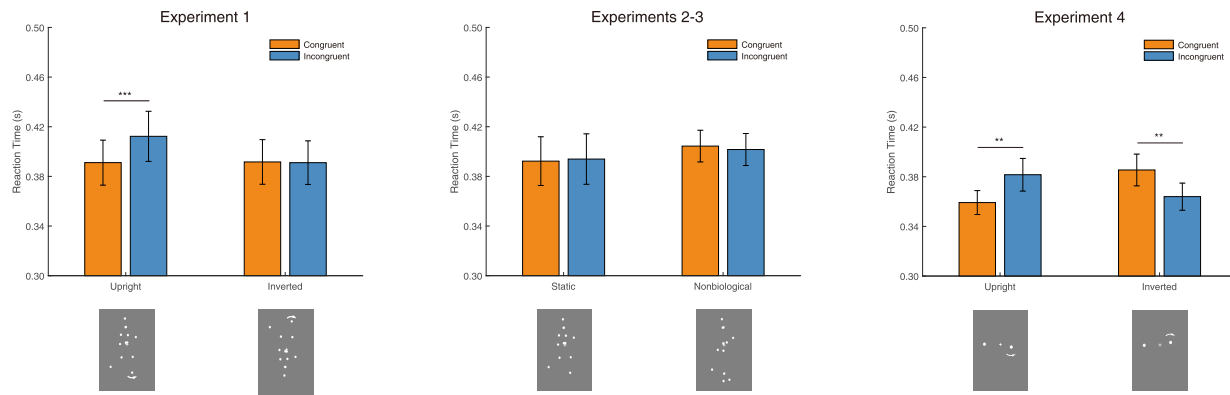


Figure 2. Results from Experiments 1 to 4. In Experiment 1, a cross-modal attentional effect was observed with the upright but not the inverted BM cues. Such an effect vanished in Experiment 2 (static BM frames) and Experiment 3 (nonbiological motion sequences). Furthermore, the observed cross-modal attentional effect could be extended to feet motion sequences (Experiment 4), which was also sensitive to the orientation of the feet motion cues. Error bars show standard errors. *** $p < 0.001$; ** $p < 0.01$.

Results

In all four experiments, trials with incorrect responses and reaction times (RTs) less than 100 ms or greater than 1500 ms were excluded from the statistical analysis (1.0% of all trials). Mean RTs from Experiments 1 through 4 are shown in Figure 2.

In Experiment 1, data for the mean RTs were entered into a 2×2 repeated-measures analysis of variance (ANOVA) with the two within-subjects factors of cue–probe congruency (congruent vs. incongruent) and BM orientation (upright vs. inverted). Results revealed a significant interaction between cue–probe congruency and BM orientation, $F(1, 23) = 9.300$, $p = 0.006$, $\eta_p^2 = 0.288$. Specifically, when upright point-light BM sequences were presented as central cues, a paired t -test showed that participants responded significantly faster to auditory targets presented in the congruent trials than those in the incongruent trials—391.1 ms vs. 412.3 ms; 95% confidence interval (CI) for mean difference, 10.3–32.0; $t(23) = 4.044$; $p < 0.001$; Cohen's $d = 0.825$; Bayes factor (BF_{10}) = 127.979—even when they were explicitly told that the walking direction of BM did not predict the location of the auditory targets. This cross-modal attentional effect, however, disappeared when inverted point-light walkers were used as central cues—391.6 ms vs. 391.1 ms; 95% CI for mean difference, –9.4 to 8.2; $t(23) = -0.139$; $p = 0.891$; Cohen's $d = 0.028$; $BF_{10} = 0.194$ —consistent with previous observations from unimodal studies (Bardi et al., 2015; Shi et al., 2010). In other words, the attentional orienting triggered by nonpredictive BM cues extended beyond the modality of vision to that of audition and critically depended on the orientation of the BM cues.

To rule out the possibility that the observed cross-modal attentional orienting effect was contributed by the viewpoint information of the point-light figures (e.g., a point-light figure facing left or right) rather than the walking direction information of BM, we employed static point-light figures as central cues in Experiment 2. A paired t -test revealed no significant difference in RTs between auditory probes presented in the facing direction and those in the opposite direction—392.2 ms vs. 393.9 ms; 95% CI for mean difference, –9.0 to 12.4; $t(23) = 0.320$; $p = 0.752$; Cohen's $d = 0.065$; $BF_{10} = 0.278$ —indicating that static frames with a discernable human figure could not produce cross-modal attentional orienting. Furthermore, to examine if the observed cross-modal attentional effect was indeed triggered by the biological characteristics of the BM signals, we conducted an additional control experiment (Experiment 3) in which nonbiological motion sequences were used as central cues. Again, results showed that the difference between the congruent and the incongruent conditions obtained from upright BM stimuli disappeared—404.4 ms vs. 401.6 ms; 95% CI for mean difference, –10.0 to 4.5; $t(23) = -0.786$; $p = 0.440$; Cohen's $d = 0.160$; $BF_{10} = 0.131$ —indicating that nonbiological motion sequences, although sharing identical moving trajectories with BM stimuli, could not elicit cross-modal attentional orienting. Collectively, these findings demonstrate that the observed cross-modal attentional effect was essentially triggered by the biological characteristics contained in the motion rather than the form signals.

To further probe the dependence of the observed cross-modal attentional effect on the global configuration of BM signals, we adopted feet motion sequences in Experiment 4. The feet motion sequences consisted of only the two point lights of the ankles

representing the walkers' feet, which obviously had no global configuration. Similar to Experiment 1, a significant interaction of congruency (congruent vs. incongruent) and feet motion orientation (upright vs. inverted) was found, $F(1, 23) = 15.631$, $p = 0.001$, $\eta_p^2 = 0.405$. Further analyses revealed that participants performed better in the congruent condition than in the incongruent condition when upright feet motion cues were presented, 359.2 ms vs. 381.6 ms; 95% CI for mean difference, 7.9–36.9; $t(23) = 3.209$; $p = 0.004$; Cohen's $d = 0.655$; $BF_{10} = 21.512$. Moreover, the magnitude of the cross-modal attentional effect induced by feet motion cues (calculated using the difference in the mean RT obtained under the incongruent condition vs. that under the congruent condition divided by their sum, $\frac{RT_{\text{incongruent}} - RT_{\text{congruent}}}{RT_{\text{incongruent}} + RT_{\text{congruent}}}$) was not different from that induced by intact BM cues, $t(46) = -0.241$, $p = 0.811$, Cohen's $d = 0.070$, $BF_{10} = 0.294$. These results suggest that the cross-modal attentional effect did not necessarily rely on the global configuration and could be induced by the local motion signals alone. Intriguingly, when the feet motion cues were shown inverted, the attentional effect was also significant but revealed a reverse pattern—385.4 ms vs. 364.0 ms; 95% CI for mean difference, -35.0 to -7.9; $t(23) = -3.278$; $p = 0.003$; Cohen's $d = 0.669$; $BF_{10} = 24.812$ —which is in line with our previous unimodal study (L. Wang et al., 2014). That is, the performance of the observers was worse when the probe was presented in the motion direction of the inverted feet motion cues (congruent condition) than in the opposite direction (incongruent condition). This pattern of result seems to arise from the translatory (extrinsic) motion in the stance phase that essentially points to the opposite direction of the walking direction (see Methods for more detail). It should be noted that the inversion of the feet motion cues disrupted only the intrinsic biological information contained in the upright feet motion cues (e.g., vertical acceleration due to muscle activity and gravity), whereas the horizontal, translatory motion in the stance phase remained unchanged. In other words, the attentional effect induced by the walking direction of the upright feet motion cues overrode the effect from the translatory motion (which was opposite to the walking direction). Taken together, these findings demonstrated that the walking direction carried by the motion of feet was effective at triggering a cross-modal attentional effect.

Discussion

Humans live in an interconnected world and always need to share attention with interactive social partners in order to better detect objects or events from different modalities. It is therefore necessary to investigate the

effect of social attention in a multisensory setting that is more representative of our daily life. Here, we examined social attention induced by BM cues under visual–auditory cross-modal conditions and found that BM cues could trigger auditory attentional orienting. The observed cross-modal social attention appeared to vanish when the BM cues were shown upside-down, reflecting an inversion effect associated with BM processing (Chang & Troje, 2009; Troje & Westhoff, 2006). Moreover, the effect was not due to the viewpoint effect of the biological figure, as a static point-light human figure showed no such attentional effect. Critically, such an effect completely disappeared when critical biological characteristics were removed. Furthermore, local BM cues, which consisted of only the two point lights of the ankles and had no global configuration, could produce similar cross-modal attentional orienting, and this effect was also sensitive to the orientation of the local cues. Together, these findings demonstrate a cross-modal attentional effect driven by biological characteristics embedded in the motion signals independent of global configuration.

The focus in the research on BM processing has traditionally been limited to the visual modality. Given the ubiquity of multisensory information in real life, it is not surprising that multisensory research has captured the attention of scientists from the research field of BM in recent years. For example, it was found that auditory motion in the same direction as the BM target could facilitate its detectability (A. Brooks et al., 2007). Moreover, auditory cues affected the extraction of high-level features (e.g., gender) from BM displays (van der Zwan et al., 2009). Here, we went a step further by employing auditory stimuli in a social attention task and demonstrated that visual BM displays could conversely modulate attention to auditory cues. It should be noted that there is a spatial mismatch between visual and auditory signals in the cross-modal social attention task. In the future, it will be important to investigate the influences of spatial mismatch on multisensory interactions of BM (Harrison, Wuerger, & Meyer, 2010).

On the other hand, our findings extended prior unimodal social attention studies and showed that BM-triggered attention could span to another sensory modality (i.e., audition). A similar cross-modal social attention has been observed with another type of social cue (i.e., eye gaze) (Doruk, Chanes, Malavera, Merabet, Valero-Cabre, & Fregni, 2018; Newport & Howarth, 2009; Pines & Bekkering, 2010; Soto-Faraco, Sinnett, Alsius, & Kingstone, 2005). It has been demonstrated that gaze cues can induce shifts in auditory or tactile attention (Newport & Howarth, 2009; Soto-Faraco et al., 2005). Exogenous attention caused by the visual peripheral nonsocial cue has also been investigated in a similar cross-modal paradigm. For example, several studies have reported that uninformative visual

peripheral cues cannot trigger attentional orienting to auditory targets (Bachtel & Butter, 1988; Spence & Driver, 1997). Combined with previous evidence, our findings obtained here reflect the specificity and cross-modal nature of social attention.

Notably, such cross-modal social attention could be extended to local BM stimuli without any global configuration. The important role of local motion in BM perception has long been overlooked, but recently some researchers investigated its ecological and social significance by adopting novel stimuli (e.g., only the two point lights of the ankles representing the feet) that were entirely devoid of configural cues and retained only local motion signals (Troje & Westhoff, 2006; L. Wang & Jiang, 2012; L. Wang, Zhang, He, & Jiang, 2010; L. Wang et al., 2014; Y. Wang et al., 2018). These studies have emphasized the special role of the motion of the feet in the perception of BM walking direction (Chang & Troje, 2009; Gurnsey, Roddy, & Troje, 2010; M. H. Johnson, 2006; Saunders, Suchan, & Troje, 2009; Troje & Westhoff, 2006). Moreover, it has been shown that feet motion cues not only can be processed independent of global configuration but also can trigger reflexive attentional effect even without an observer's explicit awareness of the biological nature of the motion (L. Wang et al., 2014). It has hence been suggested that there might exist a specialized and intrinsic brain mechanism sensitive to the direction of the limbs of another moving creature (i.e., life motion detector) (Troje & Westhoff, 2006; L. Wang & Jiang, 2012; L. Wang et al., 2014). However, all of these previous studies explored local BM processing exclusively within a unimodal visual setting. In the present study, we first examined the processing of feet motion cues under a visual–auditory cross-modal condition and showed that local BM cues could be effective at triggering shifts in auditory attention. Our findings provide evidence for the automatic processing of local BM cues from a multisensory perspective and suggest the existence of a dedicated multimodal mechanism subserving local BM processing.

It is worth noting that the present study does not allow determining whether the cross-modal social attention is driven by facilitated attentional engagement or by delayed disengagement. By comparing reaction times to the averted (congruent or incongruent) and direct (neutral) gaze condition, a previous unimodal study demonstrated that gaze-triggered social attention resulted from enhanced engagement rather than delayed disengagement (Friesen & Kingstone, 1998). In addition, a cross-modal exogenous attention study showed that the emotional valence of the peripheral visual cues could modulate the roles of engagement and disengagement components in the auditory attention (Harrison & Woodhouse, 2016). Future research should include a neutral cue (e.g., a point-light walker facing toward the viewer) to investigate whether the

cross-modal cueing effect obtained in the present study was due to facilitated engagement or delayed disengagement.

With regard to the underlying neural mechanisms, brain regions that have been previously implicated in BM perception, attentional orienting, and multisensory processing are likely involved in the obtained cross-modal social attention triggered by BM. Transcranial magnetic stimulation (Grossman, Battelli, & Pascual-Leone, 2005) and functional magnetic resonance imaging (Grossman & Blake, 2001; Grossman, Donnelly, Price, Pickens, Morgan, Neighbor, & Blake, 2000) studies have consistently shown that the superior temporal sulcus (STS), which plays a crucial role in understanding intentions and goals (Gobbini, Koralek, Bryan, Montgomery, & Haxby, 2007), is central to BM perception (Puce & Perrett, 2003). In addition, BM task performance (walking direction discrimination) is associated with activity in the bilateral STS (Herrington, Nymberg, & Schultz, 2011). Particularly, this area is proposed to subserve the reflexive shift of attention induced by biological signals (e.g., gaze and BM) in unimodal studies (Kingstone, Friesen, & Gazzaniga, 2000; L. Wang et al., 2014). Note that the STS also plays a crucial role in multisensory processing (Calvert, 2001) and, more specifically, the combined processing of visual and auditory BM signals (Meyer, Greenlee, & Wuerger, 2011; Meyer, Harrison, & Wuerger, 2013). Therefore, it is reasonable to expect that the STS is critically involved in the cross-modal orienting effect observed in our study; however, the specific neural network behind BM-mediated cross-modal social attention remains an important question worthy of further investigation.

In conclusion, the current study clearly demonstrates that BM signals independent of global configuration have consequences that reach beyond their own sensory modality, enhancing related auditory information processing. These findings lend strong support for the existence of a supramodal attention mechanism that is specifically tuned to life motion signals. Life motion signals may inform us about the potential presence of a nearby event (e.g., food) that is out of sight and thus synchronize our perceptions of the complex world.

Keywords: biological motion, cross-modal attention, social attention, vision, audition

Acknowledgments

Supported by grants from the National Natural Science Foundation of China (31525011, 31671137, and 31830037), the Strategic Priority Research Program (XDB32010300), the National Key Research and Development Project (2020AAA0105600), and

the Key Research Program of Frontier Sciences (QYZDB-SSW-SMC030) and by the Beijing Municipal Science & Technology Commission, Shenzhen-Hong Kong Institute of Brain Science, and Fundamental Research Funds for the Central Universities.

Commercial relationships: none.

Corresponding author: Li Wang.

Email: wangli@psych.ac.cn.

Address: Institute of Psychology, Chinese Academy of Sciences, Beijing, China.

References

- Arrighi, R., Marini, F., & Burr, D. (2009). Meaningful auditory information enhances perception of visual biological motion. *Journal of Vision*, 9(4):25, 1–7, <https://doi.org/10.1167/9.4.25>.
- Bardi, L., Di Giorgio, E., Lunghi, M., Troje, N. F., & Simion, F. (2015). Walking direction triggers visuo-spatial orienting in 6-month-old infants and adults: An eye tracking study. *Cognition*, 141, 112–120.
- Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.
- Bertenthal, B. I., & Pinto, J. (1994). Global processing of biological motions. *Psychological Science*, 5(4), 221–225.
- Birmingham, E., & Kingstone, A. (2009). Human social attention. *Progress in Brain Research*, 176, 309–320.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436.
- Brooks, A., Schouten, B., Troje, N. F., Verfaillie, K., Blanke, O., & van der Zwan, R. (2008). Correlated changes in perceptions of the gender and orientation of ambiguous biological motion figures. *Current Biology*, 18(17), R728–R729.
- Brooks, A., van der Zwan, R., Billard, A., Petreska, B., Clarke, S., & Blanke, O. (2007). Auditory motion affects visual biological motion processing. *Neuropsychologia*, 45(3), 523–530.
- Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, 8(6), 535–543.
- Buchtel, H. A., & Butter, C. M. (1988). Spatial attentional shifts: Implications for the role of polysensory mechanisms. *Neuropsychologia*, 26(4), 499–509.
- Calvert, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex*, 11(12), 1110–1123.
- Chang, D. H. F., & Troje, N. F. (2009). Acceleration carries the local inversion effect in biological motion perception. *Journal of Vision*, 9(1):19, 1–17, <https://doi.org/10.1167/9.1.19>.
- Di Giorgio, E., Loveland, J. L., Mayer, U., Rosa-Salva, O., Versace, E., & Vallortigara, G. (2017). Filial responses as predisposed and learned preferences: Early attachment in chicks and babies. *Behavioural Brain Research*, 325(Pt B), 90–104.
- Doruk, D., Chanes, L., Malavera, A., Merabet, L. B., Valero-Cabre, A., & Fregni, F. (2018). Cross-modal cueing effects of visuospatial attention on conscious somatosensory perception. *Heliyon*, 4(4), e00595.
- Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition*, 6(5), 509–540.
- Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, 5(3), 490–495.
- Friesen, C. K., Ristic, J., & Kingstone, A. (2004). Attentional effects of counterpredictive gaze and arrow cues. *Journal of Experimental Psychology: Human Perception and Performance*, 30(2), 319–329.
- Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, 133(4), 694–724.
- Frischen, A., Smilek, D., Eastwood, J. D., & Tipper, S. P. (2007). Inhibition of return in response to gaze cues: The roles of time course and fixation cue. *Visual Cognition*, 15(8), 881–895.
- Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., & Haxby, J. V. (2007). Two takes on the social brain: A comparison of theory of mind tasks. *Journal of Cognitive Neuroscience*, 19(11), 1803–1814.
- Grossman, E. D., Battelli, L., & Pascual-Leone, A. (2005). Repetitive TMS over posterior STS disrupts perception of biological motion. *Vision Research*, 45(22), 2847–2853.
- Grossman, E. D., & Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vision Research*, 41(10), 1475–1482.
- Grossman, E. D., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., . . . Blake, R. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience*, 12(5), 711–720.

- Gurnsey, R., Roddy, G., & Troje, N. F. (2010). Limits of peripheral direction discrimination of point-light walkers. *Journal of Vision*, 10(2):15, 1–17, <https://doi.org/10.1167/10.2.15>.
- Harrison, N. R., & Woodhouse, R. (2016). Modulation of auditory spatial attention by visual emotional cues: Differential effects of attentional engagement and disengagement for pleasant and unpleasant cues. *Cognitive Processing*, 17(2), 205–211.
- Harrison, N. R., Wuerger, S. M., & Meyer, G. F. (2010). Reaction time facilitation for horizontally moving auditory-visual stimuli. *Journal of Vision*, 10(14):16, 1–21, <https://doi.org/10.1167/10.14.16>.
- Herrington, J. D., Nymberg, C., & Schultz, R. T. (2011). Biological motion task performance predicts superior temporal sulcus activity. *Brain and Cognition*, 77(3), 372–381.
- Ji, H., Wang, L., & Jiang, Y. (2020). Cross-category adaptation of reflexive social attention. *Journal of Experimental Psychology: General*. Published online ahead of print, <https://doi.org/10.1037/xge0000766>.
- Johnson, K. L., McKay, L. S., & Pollick, F. E. (2011). He throws like a girl (but only when he's sad): Emotion affects sex-decoding of biological motion displays. *Cognition*, 119(2), 265–280.
- Johnson, M. H. (2006). Biological motion: A perceptual life detector? *Current Biology*, 16(10), R376–R377.
- Kingstone, A., Friesen, C. K., & Gazzaniga, M. S. (2000). Reflexive joint attention depends on lateralized cortical connections. *Psychological Science*, 11(2), 159–166.
- Klein, J. T., Shepherd, S. V., & Platt, M. L. (2009). Social attention and the brain. *Current Biology*, 19(20), 958–962.
- Klin, A., Lin, D. J., Gorrindo, P., Ramsay, G., & Jones, W. (2009). Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature*, 459(7244), 257–261.
- Kuhlmeier, V. A., Troje, N. F., & Lee, V. (2010). Young infants detect the direction of biological motion in point-light displays. *Infancy*, 15(1), 83–93.
- Manera, V., Schouten, B., Becchio, C., Bara, B. G., & Verfaillie, K. (2010). Inferring intentions from biological motion: A stimulus set of point-light communicative interactions. *Behavior Research Methods*, 42(1), 168–178.
- Mathews, A., Fox, E., Yiend, J., & Calder, A. (2003). The face of fear: Effects of eye gaze and emotion on visual attention. *Visual Cognition*, 10(7), 823–835.
- Mendonca, C., Santos, J. A., & Lopez-Moliner, J. (2011). The benefit of multisensory integration with biological motion signals. *Experimental Brain Research*, 213(2–3), 185–192.
- Meyer, G. F., Greenlee, M., & Wuerger, S. (2011). Interactions between auditory and visual semantic stimulus classes: Evidence for common processing networks for speech and body actions. *Journal of Cognitive Neuroscience*, 23(9), 2291–2308.
- Meyer, G. F., Harrison, N. R., & Wuerger, S. M. (2013). The time course of auditory-visual processing of speech and body actions: Evidence for the simultaneous activation of an extended neural network for semantic processing. *Neuropsychologia*, 51(9), 1716–1725.
- Newport, R., & Howarth, S. (2009). Social gaze cueing to auditory locations. *Quarterly Journal of Experimental Psychology*, 62(4), 625–634.
- Nuku, P., & Bekkering, H. (2008). Joint attention: Inferring what others perceive (and don't perceive). *Consciousness and Cognition*, 17(1), 339–349.
- Nummenmaa, L., & Calder, A. J. (2009). Neural mechanisms of social attention. *Trends in Cognitive Sciences*, 13(3), 135–143.
- Oram, M. W., & Perrett, D. I. (1994). Responses of anterior superior temporal polysensory (STPa) neurons to “biological motion” stimuli. *Journal of Cognitive Neuroscience*, 6(2), 99–116.
- Pavlova, M. A. (2012). Biological motion processing as a hallmark of social cognition. *Cerebral Cortex*, 22(5), 981–995.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.
- Pines, N., & Bekkering, H. (2010). When one sees what the other hears: Crossmodal attentional modulation for gazed and non-gazed upon auditory targets. *Consciousness and Cognition*, 19(1), 135–143.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1), 3–25.
- Puce, A., & Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 358(1431), 435–445.
- Rosa Salva, O., Mayer, U., & Vallortigara, G. (2015). Roots of a social brain: Developmental models of emerging animacy-detection mechanisms. *Neuroscience & Biobehavioral Reviews*, 50, 150–168.
- Saunders, D. R., Suchan, J., & Troje, N. F. (2009). Off on the wrong foot: Local features in biological motion. *Perception*, 38(4), 522–532.
- Schouten, B., Troje, N. F., Vroomen, J., & Verfaillie, K. (2011). The effect of looming and receding sounds on the perceived in-depth orientation of depth-ambiguous biological motion figures. *PLoS One*, 6(2), e14725.

- Shepherd, S. V. (2010). Following gaze: Gaze-following behavior as a window into social cognition. *Frontiers in Integrative Neuroscience*, 4, 5.
- Shi, J., Weng, X., He, S., & Jiang, Y. (2010). Biological motion cues trigger reflexive attentional orienting. *Cognition*, 117(3), 348–354.
- Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences, USA*, 105(2), 809–813.
- Soto-Faraco, S., Sinnett, S., Alsius, A., & Kingstone, A. (2005). Spatial orienting of tactile attention induced by social cues. *Psychonomic Bulletin & Review*, 12(6), 1024–1031.
- Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, 59(1), 1–22.
- Sweeny, T. D., Wurnitsch, N., Gopnik, A., & Whitney, D. (2013). Sensitive perception of a person's direction of walking by 4-year-old children. *Developmental Psychology*, 49(11), 2120–2124.
- Thomas, J. P., & Shiffrar, M. (2010). I can see you better if I can hear you coming: Action-consistent sounds facilitate the visual detection of human gait. *Journal of Vision*, 10(12):14, 1–11, <https://doi.org/10.1167/10.12.14>.
- 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):12, 1–7, <https://doi.org/10.1167/7.10.12>.
- Thompson, J., & Parasuraman, R. (2012). Attention, biological motion, and action recognition. *NeuroImage*, 59(1), 4–13.
- Tipples, J. (2008). Orienting to counterpredictive gaze and arrow cues. *Perception & Psychophysics*, 70(1), 77–87.
- Troje, N. F., & Westhoff, C. (2006). The inversion effect in biological motion perception: Evidence for a “life detector”? *Current Biology*, 16(8), 821–824.
- Vallortigara, G., & Regolin, L. (2006). Gravity bias in the interpretation of biological motion by inexperienced chicks. *Current Biology*, 16(8), R279–R280.
- Vallortigara, G., Regolin, L., & Marconato, F. (2005). Visually inexperienced chicks exhibit spontaneous preference for biological motion patterns. *PLoS Biology*, 3(7), e208.
- van der Zwan, R., MacHatch, C., Kozlowski, D., Troje, N. F., Blanke, O., & Brooks, A. (2009). Gender bending: Auditory cues affect visual judgements of gender in biological motion displays. *Experimental Brain Research*, 198(2), 373–382.
- Vanrie, J., & Verfaillie, K. (2004). Perception of biological motion: A stimulus set of human point-light actions. *Behavior Research Methods, Instruments, & Computers*, 36(4), 625–629.
- Wang, L., & Jiang, Y. (2012). Life motion signals lengthen perceived temporal duration. *Proceedings of the National Academy of Sciences, USA*, 109(11), 4043–4043.
- Wang, L., Wang, Y., Xu, Q., Liu, D., Ji, H., & Yu, Y., ...Jiang, Y. (2020). Heritability of reflexive social attention triggered by eye gaze and walking direction: Common and unique genetic underpinnings. *Psychological Medicine*, 50(3), 475–483.
- Wang, L., Yang, X., Shi, J., & Jiang, Y. (2014). The feet have it: Local biological motion cues trigger reflexive attentional orienting in the brain. *NeuroImage*, 84(1), 217–224.
- Wang, L., Zhang, K., He, S., & Jiang, Y. (2010). Searching for life motion signals: Visual search asymmetry in local but not global biological-motion processing. *Psychological Science*, 21(8), 1083–1089.
- Wang, Y., Wang, L., Xu, Q., Liu, D., Chen, L., & Troje, N. F., ...Jiang, Y. (2018). Heritable aspects of biological motion perception and its covariation with autistic traits. *Proceedings of the National Academy of Sciences, USA*, 115(8), 1937–1942.
- Wuerger, S. M., Crocker-Buque, A., & Meyer, G. F. (2012). Evidence for auditory-visual processing specific to biological motion. *Seeing and Perceiving*, 25(1), 15–28.
- Zhao, J., Wang, L., Wang, Y., Weng, X., Li, S., & Jiang, Y. (2014). Developmental tuning of reflexive attentional effect to biological motion cues. *Scientific Reports*, 4, 5558.