Cogn Process (2017) 18:195–203
DOI 10.1007/s10339-017-0792-y

SHORT COMMUNICATION

Attention to body-parts varies with visual preference and verb–effector associations

Ty W. Boyer[1] · Josita Maouene[2] · Nitya Sethuraman[3]

Received: 15 June 2016 / Accepted: 31 January 2017 / Published online: 9 February 2017

Abstract Theories of embodied conceptual meaning suggest fundamental relations between others' actions, language, and our own actions and visual attention processes. Prior studies have found that when people view an image of a neutral body in a scene they first look toward, in order, the head, torso, hands, and legs. Other studies show associations between action verbs and the body-effectors used in performing the action (e.g., "jump" with feet/legs; "talk" with face/head). In the present experiment, the visual attention of participants was recorded with a remote eye-tracking system while they viewed an image of an actor pantomiming an action and heard a concrete action verb. Participants manually responded whether or not the action image was a good example of the verb they heard. The eye-tracking results confirmed that participants looked at the head most, followed by the hands, and the feet least of all; however, visual attention to each of the body-parts also varied as a function of the effector associated with the spoken verb on image/verb congruent trials, particularly for verbs associated with the legs. Overall, these results suggest that language influences some perceptual processes; however, hearing auditory verbs did not alter the previously reported fundamental hierarchical sequence of directed attention, and fixations on specific body-effectors may not be essential for verb comprehension as peripheral visual cues may be sufficient to perform the task.

Keywords Embodied cognition · Action perception · Verb processing · Visual attention

Handling Editor: Katsumi Watanabe (University of Tokyo); Reviewers: Fernando Marmolejo-Ramos (Stockholm University), Kyoshiro Sasaki (Kyushu University).

✉ Ty W. Boyer
tboyer@georgiasouthern.edu

Josita Maouene
maouenej@gvsu.edu

Nitya Sethuraman
nitya@umich.edu

[1] Department of Psychology, Georgia Southern University, P.O. Box 8041, Statesboro, GA 30460-8041, USA

[2] Psychology Department, Grand Valley State University, 1 Campus Dr, AuSable Hall, Allendale, MI 49401, USA

[3] Department of Behavioral Sciences, University of Michigan-Dearborn, 4901 Evergreen Road, Dearborn, MI 48128, USA

Introduction

There is a long tradition of treating the body and its parts as special in visual cognition: Viewers preferentially attend to others' bodies, particularly their faces and eyes (Buswell 1935; Yarbus 1967), show first fixations on an ordered preference of head, torso, hand, and legs (Kano and Tomonaga 2009) and attend to body-regions when making meaningful socio-cultural and emotional inferences (Majid 2010; Nummenmaa et al. 2013). Another long research tradition, often referred to as embodied cognition, treats body morphology, action, and perception as fundamentally connected to language and semantic processing (Barsalou 2008; Clark 1997; Glenberg et al. 2013; Varela et al. 1991; Zwaan 2014). These lines of research together have been used to examine how we process multimodal and dynamic everyday scenes, for example, one person hearing "take this" while his or her eye movements focus on a particular region of the speaker's body where an object might be found (for a review see Mishra and Marmolejo-Ramos 2010). This prior research on embodiment—involving motor, linguistic, multisensory, and multimodal dynamic

representations—has focused on objects and object affordances in scene understanding. Object affordances play a special role in embodiment, drawing attention by guiding eye movements during language–scene interaction (Kamide et al. 2003; Mishra and Marmolejo-Ramos 2010), which confounds our understanding of the role of body-effectors in verb meaning.

In the present study, we examine multimodal representations, devoid of objects, to examine the dynamic processing of linguistic, auditory, visual, and attentional cues and preferences. Our strategy was to measure participants' eye movements when viewing an image of a human in action (e.g., a woman pantomiming a pouring action) simultaneous with hearing a matching or mismatching verb (e.g., "*pour*" or "*walk*", respectively). In this sense, we examine how visual attention is influenced by a general tendency to look toward specific body-regions in a particular sequence, and whether this is modulated by the body-regions associated with the verb's meaning.

Internally, our body-effectors contribute to the understanding of our own activities (Thelen and Smith 1994) and emotions (Nummenmaa et al. 2013), enable gesture-based communication (Goldin-Meadow and Beilock 2010), and are recruited for kinesthetically imagining movements from a first-person perspective (Stins et al. 2015). Externally, others' bodies capture our visual attention like no other object (de Gelder et al. 2010), allow us to infer others' intentions (Blake and Shiffrar 2007; Blakemore and Decety 2001), learn new motor skills (Brown and Robertson 2007; Urgolites and Wood 2013), and imitate movements (Chartrand and Bargh 1999; Heyes 2011). We are particularly attracted to others' faces, which provide us with social and emotional information (Cole et al. 2013; Ro et al. 2007), and others' hands, which may contribute to our understanding of actions (Papeo et al. 2010; von Hofsten 2004).

Converging evidence also suggests strong associations between body-effectors, motor movements, and linguistic systems (Arévalo et al. 2012; Mishra and Marmolejo-Ramos 2010; Pulvermüller 2005) and a meaningful sensory-motor link between the body, as represented by the motor cortex, and action words (e.g., Cardona et al. 2014; Carota et al. 2012; cf. Postle et al. 2008). The highly coherent semantic relations found through free-association data indicates early-learned verbs cluster around five main body-regions (eye/mouth/ear/leg/hand), providing additional evidence for a link between the body and language (see Maouene et al. 2008 for English verbs; Duggirala et al. 2011 for Hindi, Urdu, and Telugu; Chen and Zhu 2014 for Mandarin Chinese). English and Cantonese speakers also show slower judgments when evaluating matches between verbs and line drawings that involve the same body-effector, which has been suggested to be due to inhibitory effects from the activation of effector-specific memory circuits (Bergen et al. 2010). Even an implicit relationship between the body, action, visual attention, and language is shown by a faster arm contraction than arm expansion response to hearing "open the drawer" (Glenberg and Kaschak 2002), by the role of context in studies of language comprehension (e.g., Barsalou 2008; Kemmerer 2015), and different response times and patterns of neural activation when action sentences and responses to them are compatible versus incompatible (Aravena et al. 2010).

The primary aim of the present research is to further examine how language, motor systems, and visual attention potentially interact. Language is inherently ambiguous and imperfect, and yet we are able to come to mutual understanding and communicate with each other; however, language is not used in isolation. We make use of the wealth of preceding, simultaneous, and following auditory, visual, and correlated cues, including cues from body morphology, that are available to us. Understanding how we process language in action involves examining how multimodal cues and automatic biases compete with each other in attention. Here, we examine our automatic biases for competing visual cues (body form/position) and auditory cues (hearing a word) in interpreting action words. We are particularly interested in whether and how information about the body influences our understanding of language, by providing visual information about body morphology while participants are given auditory linguistic cues—likely one of the more frequent ways in which we encounter language usage.

In the present study, participants viewed photographs of actors pantomiming common actions, while hearing either congruent or incongruent verbs (e.g., the participant sees a picture of a person running and hears "*run*" or "*catch*"). Visual attention was measured with an eye-tracking system and analyses examined whether gaze was influenced by (a) preference for body-parts and (b) association of the action word with a body-effector. The primary hypothesis is that spoken verbs will influence participants' direction of attention, though we also predict that the influence of language will be embedded within a general tendency to attend mostly to the actors' head, followed by the hands, and then legs. Thus, the study addresses the larger question of the influence on attention of automatic biases and language-driven embodied processes.

Method

Participants

Forty-one Georgia Southern University undergraduate students (20 females, 21 males) participated for course credit. Participants self-reported handedness (34 right handed, 7 left handed) and that they had normal or

corrected to normal vision. All participants were naïve to the purpose of the research. Four participants included in the behavioral data analyses were excluded from eye-tracking analyses, one participant due to technical error and three participants due to > 50% invalid gaze data samples.

Stimuli and apparatus

Visual stimuli were photographs of six student models (four female, two male[1]) pantomiming 12 actions previously associated with three primary body-regions (Maouene et al. 2008): head/face ("find," "listen," "read," "talk"); arms/hands ("catch," "pour," "push," "throw"); legs/feet ("jump," "kick," "run," "walk"). We utilized three instances of each action.[2] The images were digitally edited to remove all background features and standardize resolution, size, and centered placement on a 500 × 700 pixel canvas (∼12.9° × 16.0° visual angle). Auditory stimuli were produced by recording one male and one female uttering each of the 12 listed verbs, with neutral prosody and tone.

All experimental variables were manipulated within-subjects with repeated presentation of each stimulus image and audio utterance, and visual and auditory stimuli were paired to create congruent and incongruent trials (see Fig. 1). Congruent trials involved action images paired with the corresponding spoken verb (e.g., an image of pouring paired with the spoken verb "*pour*"). Incongruent trials involved action images with an auditory verb associated with a different action as well as a different body-effector (e.g., an image of the head/face action "talk" paired with the legs/feet verb "*run,*" but never an image of "talk" paired with the verbs "read," "listen," or "find"). Both the image and audio stimuli were repeated across trials; specifically, each image appeared in four separate trials, once with the male and once with the female uttering the congruent verb, and once with the male and once with the female uttering an incongruent verb (i.e., a verb associated with a different body-effector). Each verb utterance was heard six times, once with each of the three congruent image instances, and three times with an incongruent body-effector action image. Thus, participants viewed images of the three model instances demonstrating each of the 12 actions, paired with either a male or female voice saying a verb that was congruent or incongruent with the image action, for a total of 144 trials (3 models × 12 actions × 2 voice genders × 2 congruence types).

Participants viewed the stimuli on the 512 × 285 mm flat-screen LCD monitor of a Tobii TX-300 eye-tracking system, presented at 1680 × 1050 pixels resolution. The system recorded a timestamp, metrics of sample validity, trial-specific information, and screen-localized *x*- and *y*-coordinates of participants' point of gaze, pupil-to-screen distance estimates, and pupil diameter at 300 Hz. Participants responded using a wireless keyboard held on their laps. E-Prime software (Psychology Software Tools Inc., Pittsburg, PA, USA) managed visual and auditory stimulus presentation, eye-tracker functionality, and response measurement.

Procedure

Following a nine-point on-screen eye-tracking system calibration routine, each participant read:

> In this experiment you will be making perceptual judgments.
> First you will see a small cross in the middle of the screen; fixate the cross.
> Next, a photo of a person will appear, and you will hear a spoken word.
> Your task is to report whether or not the photo is a good example of the spoken word.
> If the photo is a good example, press "c".
> If the photo is NOT a good example, press "m".

Each trial presented a fixation cross (1000 ms), followed by the action image simultaneous with the auditory verb. Images remained on screen for 5000 ms, and participants could respond any time during the presentation window (i.e., participants were not pressured to respond quickly). Response key was counterbalanced across participants.

Results

Behavioral Responses

Match versus Mismatch Responses

The behavioral response data indicate participants identified 94.4% of the congruent verb-image pairs as good matches (and 5.6% as poor matches) and 2.4% of incongruent verb-image pairs as good matches (and 97.6% as

[1] Four female and two male student models were used out of convenience and were recruited from the full group of student researchers who were assisting with research within the laboratory when the study was being conducted.

[2] Rather than systematically control or exhaust the adopted model-action image instances, we examined the entire library of all 72 images (i.e., 6 models × 12 actions) and selected three instances that seemed to provide the most representative exemplars of each action. As a result, for some of the actions the three instances consisted of the two male models and one of the four female models (pour, kick, walk), one of the two male models and two of the four female models (catch, push, throw, jump, find, listen, read, and talk), or three of the four female models (run), resulting in an employed stimulus set of 14 photographs of male models and 22 photographs of female models.

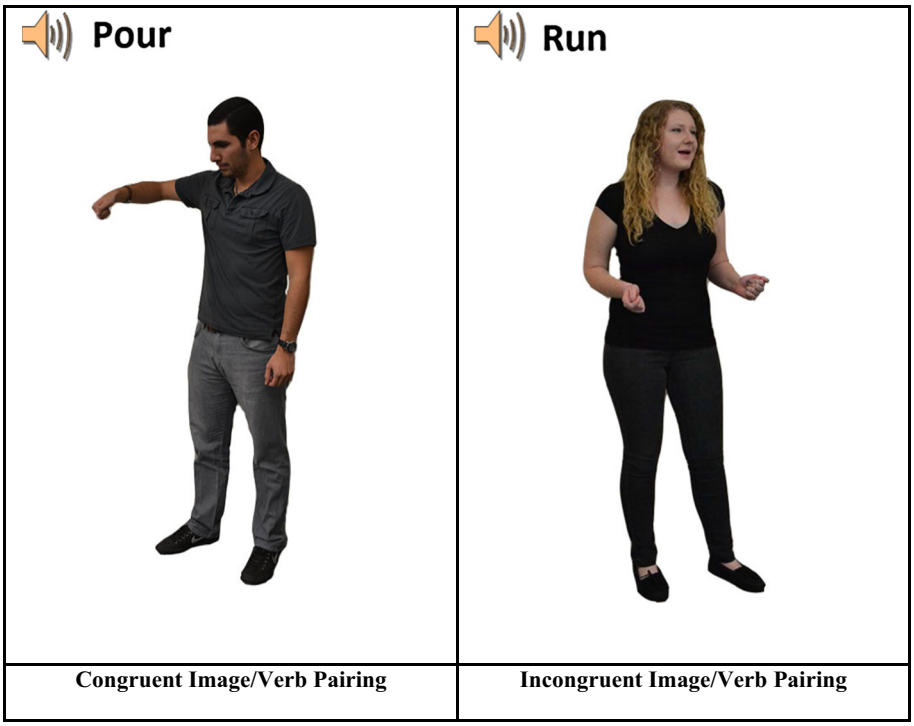

Fig. 1 Example stimuli used in the experiment. *Note*: The text presented within the image was not visible during the trial, but, rather, was presented audibly, and is included here for presentation purposes only

poor matches), representing a significant difference between good match and poor match responses to congruent and incongruent verb-image pairs, χ^2 (1, 40) = 4947.32, $p < .0001$. Characterizing congruent pairs that were identified as good matches and incongruent pairs that were identified as poor matches as accurate responses (see Table 1), there was an overall significant difference in accuracy as a function of congruence, $t(40) = 4.60$, $p < .001$, Cohen's $d = 0.99$, with a higher proportion of poor match responses to incongruent verb-image pairs ($M = 0.976$, SEM = 0.004) than good match responses to congruent verb-image pairs ($M = 0.944$, SEM = 0.006). A 3 × 3 ANOVA with image-effector (head/hands/feet) and verb-effector (head/hands/feet) entered as within-subjects factors and proportion of match responses entered as the dependent variable revealed a significant effect for verb-effector, $F(1.73, 69.05) = 11.24$, $p < .001$, $\eta_p^2 = .219$, a marginally significant effect for image-effector, $F(2, 80) = 2.98$, $p = .057$, $\eta_p^2 = .069$, and a significant verb- and image-effector interaction, $F(2.71, 108.29) = 7.58$, $p < .001$, $\eta_p^2 = .159$ (with Greenhouse-Geisser corrections where necessary, as determined by Mauchly's test of sphericity $p \leq .05$). Closer examination reveals these patterns are primarily due to head-effector images, which were less likely to be judged as good matches with congruent verbs ($M = 92.0\%$) yet more likely to be judged as poor matches with incongruent verbs, particularly leg-effector verbs ($M = 99.6\%$). In summary, participants' judgments were highly consistent with our image/verb pair classifications, suggesting that the visual stimuli were good exemplars of the intended verbs (and, by extension, the selected actions), albeit with some variability as a function of verb-image pairing and effector region.

Response Times

Overall, participants responded slightly faster for incongruent ($M = 1703.5$ ms, SEM = 38.3) versus congruent ($M = 1730.1$, SEM = 44.8) verb-image trials (see Table 2); though this RT difference is not statistically significant, $t(40) = 1.10$, $p = .28$, Cohen's $d = 0.10$, the direction of the difference indicates that there was not a speed-accuracy trade-off (i.e., responses were both more accurate and faster for incongruent trials). We also conducted a 3 × 3 ANOVA with image-effector (head, hands, feet) and verb-effector (head, hands, feet) entered as within-subjects factors and RTs as the dependent variable. This revealed significant effects for verb-effector, $F(2, 80) = 8.63$, $p < .001$, $\eta_p^2 = .178$, image-effector, $F(2, 80) = 33.66$, $p < .001$, $\eta_p^2 = .457$, and a verb- and image-effector interaction, $F(4, 160) = 7.54$, $p < .001$, $\eta_p^2 = .159$.

To further examine this three-way interaction, we conducted three separate one-way ANOVA for each body-effector verb, with the RT for each image body-effector as the dependent variables. The ANOVA for head/face verbs revealed an effect for image-effector, $F(1.74, 69.4) = 18.99$, $p < .001$, $\eta_p^2 = .322$, due to faster RTs

Table 1 Grand mean proportions of responses that congruent images represented good examples for the spoken verbs (**bold**) and incongruent images were poor examples of spoken verbs, per verb- and image-effector

Image-effector	Verb-effector		
	Head	Hands	Feet
Head	**0.920**	0.961	0.996
Hands	0.957	**0.947**	0.980
Feet	0.973	0.988	**0.965**

Table 2 Grand mean response times (in ms) per verb- and image-effector (congruent verb-image pairs in **bold**)

Image-effector	Verb-effector		
	Head	Hands	Feet
Head	**1892.6**	1808.5	1639.6
Hands	1869.2	**1738.0**	1761.3
Feet	1634.9	1576.9	**1685.2**

when head/face verbs were paired with legs/feet action images than when paired with head/face or hands images (both $p < .001$), with no pairwise RT difference between head/face and hands images ($p = .814$). The ANOVA for hands verbs also revealed an effect for image-effector, $F(2, 80) = 14.54$, $p < .001$, $\eta_p^2 = .267$, which, as was the case for head verbs, was due to faster RTs when hands verbs were paired with legs/feet action images than when paired with either head/face or hands action images (both $p \leq .003$), with less difference between head/face and hands action images ($p = .035$). Together, these analyses indicate that RTs were especially fast when the head/face and hands verbs were paired with images depicting legs/feet actions. The ANOVA for legs/feet verbs also revealed an effect for image-effector, $F(2, 80) = 8.07$, $p = .001$, $\eta_p^2 = .168$, but, in contrast with the previous analyses, was due to slower RTs when legs/feet verbs were paired with hands action images than when paired with either head/face or legs/feet action images (both $p < .001$), with no difference between head/face and legs/feet images ($p = .481$).

Eye-Tracking Data Analyses

The eye-tracking data were preprocessed to ensure sample reliability, eliminating trials with $\geq$50% invalid gaze samples (7.5% of all trials). Elimination rates did not differ as a function of verb-effector, image-effector, or verb-image interaction, all $F \leq 1.75$, $p \geq 0.18$. A customized linear interpolation algorithm replaced invalid and missing gaze samples and averaged the binocular coordinates into a single gaze position estimate.[3] Areas of interest (AOI) were identified as the mean of three independent raters' x- and y-coordinate estimates of the center of each of the five effector regions (i.e., head, hands, and feet) on each of the 36 stimulus images (see Fig. 2). Each AOI was defined as a circle with a 58-pixel radius ($\sim$1.5° visual angle), as the smallest AOI accommodating the largest head-, hand-, and foot-effectors.

First AOI

To examine visual attentional preferences, we calculated the proportion of trials on which participants first looked to each AOI (head/hands/feet), before attending to other AOIs (see Fig. 3). A 3 × 3 × 2 repeated measures ANOVA with verb-effector, AOI, and verb-image congruence as factors revealed a significant main effect of AOI, $F(1.54, 55.30) = 18.84$, $p < .001$, $\eta_p^2 = .344$, due to a higher likelihood of first looks to the head/face ($M = 0.457$ trials) than hands ($M = 0.324$) or feet ($M = 0.208$). A verb-effector by AOI interaction, $F(4, 144) = 5.49$, $p < .001$, $\eta_p^2 = .132$, and a three-way congruence by verb-effector by AOI interaction, $F(4, 144) = 28.83$, $p < .001$, $\eta_p^2 = .445$, also emerged as significant. Sidak-controlled comparisons suggest an especially lower likelihood of first looking at the feet compared with the other effectors on congruent head verb-image pairs, and an especially higher likelihood of first looking at the head on both congruent hand verb-image pairs and incongruent head verb-image pairs, all $p < .0001$.

Total Duration at Each AOI

Examining whether the body-effector associated with the spoken verbs influenced participants' cumulative visual attention (Fig. 4), a 3 × 3 × 2 repeated measures ANOVA with total gaze duration at each AOI category (head/hands/feet), per verb-effector (head/hands/legs), and verb-image congruence revealed a significant main effect of AOI, $F(1.19, 43.12) = 36.83$, $p < .001$, $\eta_p^2 = .506$, with more attention to the head ($M = 1264.0$ ms) than hands ($M = 928.6$) or feet ($M = 257.7$). The analysis also indicated a main effect of verb-effector, $F(2, 72) = 13.54$, $p < .001$, $\eta_p^2 = .273$, due to decreased attention across foot-verbs ($M = 779.5$ ms) than hand- ($M = 827.1$) or head-verbs ($M = 843.7$), and a main effect of verb-image congruence, $F(1, 36) = 6.21$, $p = .017$, $\eta_p^2 = .147$, with more looking at AOIs during congruent ($M = 829.6$) than

[3] The employed algorithm replaced each string of invalid or missing gaze samples with a stepwise average of the nearest preceding and following valid gaze samples. For instance, a sequence of estimated gaze coordinates 500, a, b, c, 520, where a, b, and c are missing or invalid samples, would be transformed to 500, 505, 510, 515, 520.

Fig. 2 Schematic of the AOIs on a stimulus image, with the center of each AOI indicated by a red dot and the area indicated by a red circle (both included here for illustrative purposes)

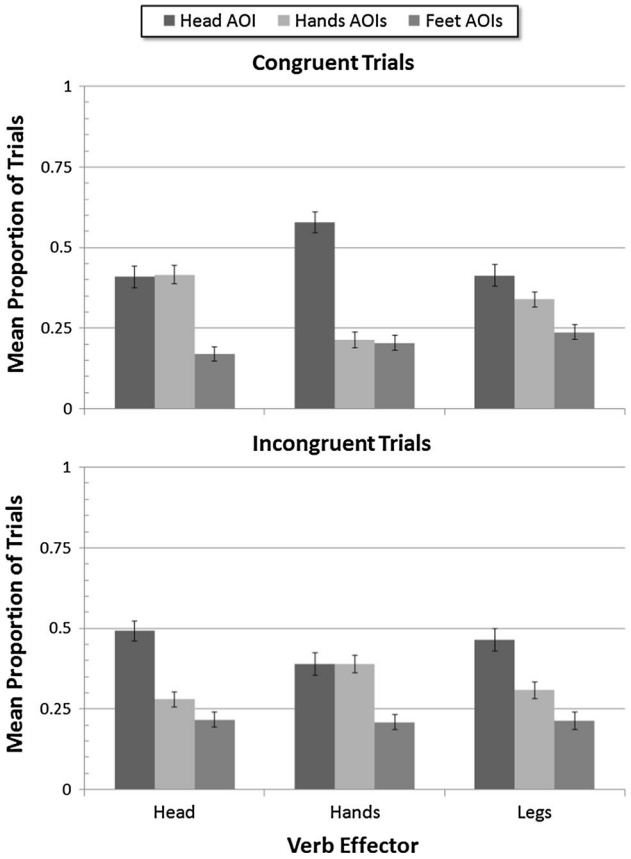

Fig. 3 Mean proportion of trials where each AOI was the first gazed region, as a function of verb-effector and verb-image congruence. Error bars represent ± standard errors of each mean

incongruent ($M = 804.0$) trials. A congruence by AOI interaction, $F(2, 72) = 4.13$, $p = .020$, $\eta_p^2 = .103$, a verb-effector by AOI interaction, $F(2.59, 93.36) = 21.24$, $p < .001$, $\eta_p^2 = .371$, and a three-way congruence by verb-effector by AOI interaction, $F(2.99, 107.63) = 9.83$, $p < .001$, $\eta_p^2 = .214$, also emerged as significant, and the congruence by verb-effector interaction was marginally significant, $F(1.50, 53.84) = 2.92$, $p = .077$, $\eta_p^2 = .075$. The most parsimonious interpretation supported by Sidak-controlled comparisons is that on congruent (but not incongruent) verb-image trials, participants looked at the body-effectors associated with the spoken verb; that is, more looking at the head for a head-verb than a hands- or feet-verb, both $p \leq .001$; at the feet for a leg-verb than a head- or hands-verb, both $p \leq .001$; and at the hands for a hands-verb than a legs-verb, $p = .001$. Planned pairwise comparisons indicated that participants looked at each respective effector longer when the verb was congruent (e.g., looked at the head longer for a head verb-image pairing than for a hand or foot verb-image pairing), all $t(36) \geq 2.07$, $p \leq .046$.[4] By comparison, look duration did not significantly differ as a function of verb-effector for incongruent trials, all $t(36) \leq 1.97$, all $p \geq .057$.[5] Essentially, participants' looking at a given effector increased specifically when the verb and image were both associated with the same effector.

Discussion

Participants viewed images of actors performing canonical pantomimed actions while hearing congruent or incongruent spoken verbs. We examined how participants' visual attention to the presented body form, devoid of objects, is influenced by both (1) visual tendencies and (2) a search across body-regions to interpret the meaning of the auditory verb. We discuss three major findings from the study that shed light on the influence of automatic biases and language-driven embodied processes on attention.

[4] Indeed, five of the six pairwise comparisons indicated $t(36) \geq 3.93$, $p < .001$, with only the comparison between looking at the hands for congruent hand verb-image pairs versus looking at the hands for congruent head verb-image pairs differing at the noted $p = .046$.

[5] Similar to footnote 1, five of the six pairwise comparisons indicated $t(36) \leq 1.59$, $p \geq .12$, with only the comparison between the looking at the hands for incongruent hand verb-image pairs versus looking at the hands for incongruent head verb-image pairs differing at the noted $p = .057$.

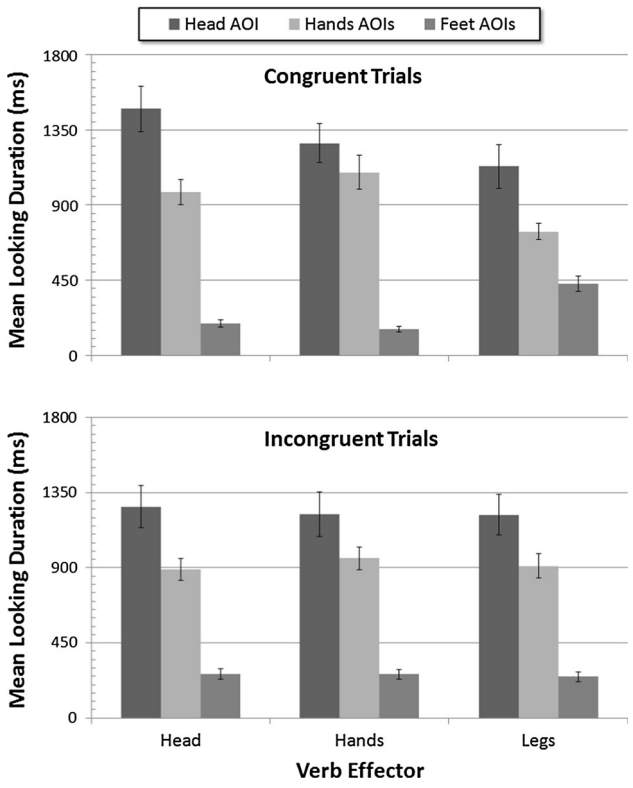

Fig. 4 Mean looking duration at each AOI as a function of the verb-effector and verb-image pair congruence. *Error bars* represent ± standard errors of each mean

First, we found that participants attended most to the head and the hands: Specifically, participants most often looked first and for the longest amount of time at the head; second most often and next longest at the hands; and last and least overall at the feet. Our results are in line with the large literature that shows faces are detected, discriminated, and recognized faster than other objects (Haxby et al. 2000; Jacques and Rossion 2006) and play a privileged role in human perceptual processes (e.g., Cole et al. 2013; de Gelder et al. 2010; Peelen and Downing 2007; Ro et al. 2007). The hands have also been found to be particularly visually attractive and informative, and facilitating of action understanding (Papeo et al. 2010; von Hofsten 2004). Participants attended least to the feet, corresponding to the small number of studies which have examined others' legs and feet as a source of socially valuable information (but see Wheaton et al. 2004; Saxe et al. 2006). Our findings may allow generalization of the "expertise effect" for face detection (Hershler and Hochstein 2005) and the proposal that we are particularly accustomed to scanning faces for available socio-emotional information (de Gelder et al. 2010). Thus, body-part viewing order may be independent of action, as we found that most of our participants' visual attention was directed to the head, regardless of the action and verb-image congruence. This is consistent with a comparative study of humans and chimpanzees viewing models showing poses instead of performing action demonstrations with no accompanying verb labels (Kano and Tomonaga 2009).

Second, we found that the default order in visual attention was influenced by language: Despite no instructions on how or where to look at the action images, participants varied how long they attended to each AOI (head/hands/feet), and how often they looked at each AOI first, as a function of verb label. Specifically, visual attention varied as a function of the body-effector associated with a spoken verb, a result mainly driven by increased attention to the leg region with a congruent verb. In other words, cues from language influence attention in modulating the default perceptual order. This suggests links between body-effectors, motor movements, and linguistic interpretation systems (Arévalo et al. 2012; Bergen et al. 2010; Desai et al. 2010; Stins et al. 2015). As reviewed, theorists suggest deep connections between the body and semantic processing (e.g., Barsalou 2008; Thelen et al. 2001; Varela et al. 1991; Zwaan 2014), and the present results provide additional support for associations between body-regions, actions, and the semantics of English verbs (Maouene et al. 2008).

Finally, the response times and cumulative-looking duration data showed that additional scanning time was recruited in the congruent trials. Specifically, we found that visual attention to body-regions increased when the body-effector associated with the spoken verb was also the action-effector in the image, particularly the leg-effector. Participants may have determined the congruence of actions associated with the feet by attending to the pictures without fixating (Posner 1980) or by using peripheral vision to the lower extremities during directed attention to the head, torso, and hands, and increased looking times to the feet to bring new information to attention (LaBerge and Brown 1989). Indeed, this pattern suggests that participants were directing attention to confirm involvement of the body-effector implied by the verb, which suggests, first, that language may be particularly effective for shifting attention to peripheral and less salient areas (Mishra 2015; Mishra and Marmolejo-Ramos 2010), and, second, that this elicits deeper processing (Bergen et al. 2010), possibly related to kinesthetic imagery (Stins et al. 2015). That this pattern did not apply similarly to incongruent trials suggests that identifying a mismatch might be accomplished via peripheral attention and shallower processing.

In summary, these results suggest that the fundamental body-part attention-based order previously documented for neutral poses applied to images of models performing actions, though this is modulated by the spoken verb and the congruency between verb and image-effectors. In this sense, language can potentially contribute to circumvention

of automatic perceptual tendencies. Thus, these results suggest a division of labor at differing levels, between shallower attentional preferences and deeper semantic processing, for confirming or disconfirming the labeling of others' actions in our task. This study also provides support for the argument for a greater focus on the body, including body morphology, in our understanding of how language and attention work together in interpretation.

References

Aravena P, Hurtado E, Riveros R, Cardona JF, Manes F, Ibáñez A (2010) Applauding with closed hands: neural signature of action-sentence compatibility effects. PLoS ONE 5:1–14

Arévalo AL, Baldo JV, Dronkers NF (2012) What do brain lesions tell us about theories of embodied semantics and the human mirror neuron system? Cortex 48:242–254

Barsalou LW (2008) Grounded cognition. Ann Rev Psychol 59:617–645

Bergen B, Lau TTC, Narayan S, Stojanovic D, Wheeler K (2010) Body part representations in verbal semantics. Mem Cognit 38:969–981

Blake R, Shiffrar M (2007) Perception of human motion. Annu Rev Psychol 58:47–73

Blakemore S-J, Decety J (2001) From the perception of action to the understanding of intention. Nat Rev Neurosci 2(8):561–567

Brown R, Robertson EM (2007) Inducing motor skill improvements with a declarative task. Nat Neurosci 10:148–149

Buswell GT (1935) How people look at pictures: a study of the psychology of perception in art. University of Chicago Press, Chicago

Cardona JF, Kergieman L, Sinay V, Gershanik O, Gelormini C, Amoruso L, Roca M, Pineda D, Trujillo N, Michon M, García AM, Szenkman D, Bekinschtein T, Manes F, Ibáñez A (2014) How embodied is action language? neurological evidence from motor diseases. Cognition 131:311–322

Carota F, Moseley R, Pulvermüller F (2012) Body-part-specific representations of semantic noun categories. J Cognit Neurosci 24:1492–1509

Chartrand TL, Bargh JA (1999) The chameleon effect: the perception-behavior link and social interaction. J Personal Soc Psychol 76:893–910

Chen Y, Zhu L (2014) Associations of body parts and early-learned Mandarin verbs and their effect on age of acquisition of these verbs. Acta Psychol Sinica 46:912–921

Clark A (1997) Being there: putting brain, body, and the world together again. MIT Press, Cambridge

Cole S, Balcetis E, Dunning D (2013) Affective signals of threat produce perceived proximity. Affective signals of threat produce perceived proximity. Psychol Sci 24:34–40

de Gelder B, Van den Stock J, Meeren HKM, Sinke CBA, Kret ME, Tamietto M (2010) Standing up for the body. Recent progress in uncovering the networks involved in the perception of bodies and bodily expressions. Neurosci Behav Rev 34:513–527

Desai RH, Binder JR, Conant LL, Seidenberg MS (2010) Activation of sensory-motor areas in sentence comprehension. Cereb Cortex 20:468–478

Duggirala V, Viswanatha N, Bapi R, Jigar P, Alladi S, Jala S, Richa N (2011) Action verbs and body parts. Int J Mind Brain Cognit 2:29–45

Glenberg AM, Kaschak MP (2002) Grounding language in action. Psychon Bull Rev 9:558–565

Glenberg AM, Witt JK, Metcalfe J (2013) From the revolution to embodiment: 25 years of cognitive psychology. Perspect Psychol Sci 8:573–585

Goldin-Meadow S, Beilock SL (2010) Action's influence on thought: the case of gesture. Psychol Sci 5(6):664–674

Haxby JV, Hoffman EA, Gobbini MI (2000) The distributed human neural system for face perception. Trends Cognit Sci 4:223–233

Hershler O, Hochstein S (2005) At first sight: a high-level pop out effect for faces. Vis Res 45:1707–1724

Heyes C (2011) Automatic imitation. Psychol Bull 137:463–483

Jacques C, Rossion B (2006) The speed of individual face categorization. Psychol Sci 17:485–492

Kamide Y, Altmann G, Haywood S (2003) Prediction and thematic information in incremental sentence processing: evidence from anticipatory eye movements. J Mem Lang 49:133–156

Kano F, Tomonaga M (2009) How chimpanzees look at pictures: a comparative eye-tracking study. Proc R Soc Lond Biol Sci 276:1949–1955

Kemmerer D (2015) Visual and motor features of the meanings of action verbs: A cognitive neuroscience perspective. In: de Almeida RG, Manouilidou C (eds) Cognitive science perspectives on verb representations and processing. Springer, Berlin, pp 189–212

LaBerge D, Brown V (1989) Theory of attentional operations in shape identification. Psychol Rev 96:101–124

Majid A (2010) Words for parts of the body. In: Malt BC, Wolff P (eds) Words and the mind: how words capture human experience. Oxford University Press, New York, pp 58–71

Maouene J, Hidaka S, Smith BL (2008) Body parts and early-learned verbs. Cognit Sci 32:1200–1216

Mishra RK (2015) Interactions between attention and language systems. A cognitive science perspective. Springer, New York

Mishra RK, Marmolejo-Ramos F (2010) On the mental representations originating during the interaction between language and vision. Cognit Process 11:295–305

Nummenmaa L, Glerean E, Hari R, Hietanen JK (2013) Bodily maps of emotions. Proc Natl Acad Sci 11:646–651

Papeo L, Negri GAL, Zadini A, Rumiati RI (2010) Action performance and action-word understanding: evidence of double dissociations in left-damaged patients. Cognit Neuropsychol 27:428–461

Peelen MV, Downing PE (2007) The neural basis of visual body perception. Nat Rev Neurosci 8:636–648

Posner MI (1980) Orienting of attention. Q J Exp Psychol 32:3–25

Postle N, McMahon KL, Ashton R, Meredith M, de Zubicaray GI (2008) Action word meaning representations in cytoarchitectonically defined primary and premotor cortices. NeuroImage 43:634–644

Pulvermüller F (2005) Brain mechanisms linking language and action. Nat Rev Neurosci 6:576–582

Ro T, Friegel A, Lavie N (2007) Attentional biases for faces and body parts. Vis Cognit 15:322–348

Saxe R, Jamal N, Powell L (2006) My body or yours? The effect of visual perspective on cortical body representations. Cereb Cortex 16:178–182

Stins JF, Schneider IK, Koole SL, Beek PJ (2015) The influence of motor imagery on postural sway: differential effects of type of body movement and person perspective. Adv Cognit Psychol 11:77–83

Thelen E, Smith LB (1994) A dynamic systems approach to the development of cognition and action. MIT Press, Cambridge

Thelen E, Schöner G, Scheier C, Smith LB (2001) The dynamics of embodiment: a field theory of infant perseverative reaching. Behav Brain Sci 24:1–86

Urgolites ZJ, Wood JN (2013) Visual long-term memory stores high-fidelity representations of observed actions. Psychol Sci 24:403–411

Varela F, Thompson E, Rosch E (1991) The embodied mind: cognitive science and human experience. MIT Press, Cambridge

von Hofsten C (2004) An action perspective on motor development. Trends Cognit Sci 8:266–272

Wheaton KJ, Thompson JC, Syngeniotis A, Abbott DF, Puce A (2004) Viewing the motion of human body parts activates different regions of premotor, temporal, and parietal cortex. NeuroImage 22:277–288

Yarbus AL (1967) Eye movements and vision. Plenum Press, New York

Zwaan R (2014) Embodiment and language comprehension: reframing the discussion. Trends Cognit Sci 18:229–234