Motor activation during action perception depends on action interpretation

Barbara Pomiechowska[1] & Gergely Csibra[1,2]

[1] Cognitive Development Center, Department of Cognitive Science, Central European

University, Budapest, Hungary

[2] Department of Psychological Sciences, Birkbeck, University of London, London, UK

Corresponding author:

Barbara Pomiechowska

Cognitive Development Center

Central European University

Nádor u. 9.

Budapest 1051, Hungary

Email: barbara.pomiechowska@gmail.com

Abstract

Since the discovery of motor mirroring, the involvement of the motor system in action interpretation has been widely discussed. While some theories proposed that motor mirroring underlies human action understanding, others suggested that it is a corollary of action interpretation. We put these two accounts to the test by employing superficially similar actions that invite radically different interpretations of the underlying intentions. Using an action-observation task, we assessed motor activation (as indexed by the suppression of the EEG mu rhythm) in response to actions typically interpreted as instrumental (e.g., grasping) or referential (e.g., pointing) towards an object. Only the observation of instrumental actions resulted in enhanced mu suppression. In addition, the exposure to grasping actions failed to elicit mu suppression when they were preceded by speech, suggesting that the presence of communicative signals modulated the interpretation of the observed actions. These results suggest that the involvement of sensorimotor cortices during action processing is conditional on a particular (instrumental) action interpretation, and that action interpretation relies on inferential processes and top-down mechanisms that are implemented outside of the motor system.

Highlights:

- The observation of grasping, but not of pointing, activates the motor system.

- Action interpretation is modulated by the presence of communicative signals.

- Motor mirroring is the consequence, rather than the source, of action interpretation.

1. Introduction

Since the seminal reports that neurons in the macaque premotor cortex are active not only when the animal executes certain movements, but also when it observes these movements executed by someone else (di Pellegrino et al., 1992; Gallese et al., 1996), a range of evidence has demonstrated that similar phenomena of action mirroring occur also in other animal species (e.g., Keller & Hahnloser, 2009; Prather et al., 2008), including humans (e.g., Buccino et al., 2004; Grèzes et al., 2003, for human infants, see Southgate et al., 2009). A widely accepted theory of the cognitive function of motor mirroring proposes that this mechanism enables observers to understand actions performed by other individuals (mirror-based account of action understanding, for a review, see Rizzolatti & Sinigaglia, 2011). Importantly, according to this account, goal attribution is achieved by the observer through simulation of witnessed actions in her own motor system, without engaging in any inferential processes (i.e., direct matching, Rizolatti et al., 2001; Gallese et al., 2004). That is, action observation triggers in the brain of the observer motor programs similar to those generated when she executes her own actions, whose goals she has direct access to and can thus use this knowledge to extrapolate the goals that guide the behavior of others.

Alternatively, it has been proposed that the involvement of the motor system in action processing might not necessarily be due to the fact that motor mirroring provides bottom-up support for action understanding. Rather, mirroring might be generated in a top-down manner as a result of action interpretation, reflecting processes involved in the reconstruction and tracking of how ongoing actions unfold (action-reconstruction account of action understanding, Csibra, 2007; see also Hickok, 2013; Jacob, 2008; Prinz, 2006) or in the preparation of responses to the actions of other social agents (social responding theory of motor mirroring, Hamilton, 2013; Sartori et al., 2013). To date, empirical evidence that could discriminate between these two theoretical approaches remains scarce despite the relevance

of this controversy to our understanding of action processing mechanisms and social cognition.

In order to determine whether motor mirroring is a prerequisite or an outcome of action interpretation it might be helpful to distinguish between instrumental and communicative actions. Such a distinction has proven useful to account for sophisticated action interpretation abilities documented in preverbal human infants (for reviews, see Csibra, 2003; Woodward, 2009), and, more specifically, to explain how the understanding of others' actions might be possible without the attribution of mental states (beliefs, intentions, desires). It was suggested that human infants are equipped with two functionally distinct action interpretation systems, instrumental and referential (Csibra, 2003), allowing them to make sense of witnessed behaviors without taking a general-purpose intentional stance (Dennett, 1987). These two systems represent two interpretational strategies that are designed to fulfill different functions: interpreting goal-directed actions, on one hand, and interpreting communication, on the other hand (Csibra, 2003; for a different version, see Gergely & Jacob, 2012). Each of these systems is characterized by a specific set of representations and computations, which result in radically different conceptualizations of witnessed events and support different kinds of inferences.

First, the teleological action interpretation system enables human infants to construe observed actions as instrumental in bringing about certain goal states (e.g., to retrieve a ball) and to draw a number of inferences that are made possible by such understanding of witnessed events (Gergely & Csibra, 2003). Specifically, infants are able to evaluate which of the actions that are available to the agent is the most efficient means to achieve its goal (e.g., which path is the most direct to reach a particular location, Csibra et al., 1999; Gergely et al., 1995; Skerry et al., 2013) and they expect agents to select such efficient means to achieve their goals (e.g., not to make unnecessary detours while moving toward a location, Csibra,

2008). Infants also succeed at inferring goals before the observed actions are completed (e.g., Csibra et al., 2003; Daum et al., 2008; Southgate & Csibra, 2009; Wagner & Carey, 2005).

Second, the referential action interpretation system biases observers to construe witnessed behaviors as highlighting those aspects of the world that are linked to the contents of a communicated message. The result of such interpretation is a shift of attentional focus from the action itself toward objects or events singled out by the action (e.g., through pointing or a gaze shift) and selected by the actor to become the deictic referent of a communicative exchange. Infants expect actions interpreted as referential to be object-directed (Csibra & Volein, 2008; Gredebäck et al., 2010), and to be consistent with each other when concurrently produced by the same person (Gliga & Csibra, 2009).

Teleological and referential action interpretation systems operate independently from each other and have different criteria of application, whose satisfaction, and thus the activation of a particular action interpretation at any given time, depends on the applicability of certain inferential schemas to the witnessed behaviors (Gergely & Csibra, 2003). Judging the observed action as being an efficient means towards its goal is consistent with the teleological stance, while the presence of ostensive signals that indicate communicative intentions of the actor (Csibra, 2010) elicits the expectation for referential actions. Besides taking rich inferential routes that start with the satisfaction of generic input conditions of the two action interpretation systems, action understanding can also be achieved by relying on the strong associative links that are formed between frequently occurring actions and their effects (Csibra & Gergely, 2007; Elsner, 2007). This applies, for example, in the case of conventionalized gestures, such as pointing, which are immediately interpreted as communicative in the absence of further evidence.

Interestingly, the competing accounts of motor mirroring described above make opposite predictions for the interpretation of such communicative behaviors. Theories of

mirror-based action understanding claim that action interpretation is driven by action simulation. Therefore, the observer's motor system should be recruited for all kinds of actions that are part of her motor repertoire. In contrast, theories denying a causal role of motor mirroring in action understanding should allow more nuanced predictions. For example, it should be predicted that the involvement of the observer's motor system might differ depending on the interpretation of the observed action (e.g., whether it is instrumental or communicative, according to the action-reconstruction framework, Csibra, 2007; whether it requires a response or not, according to the social responding framework, Hamilton, 2013).

In particular, the action-reconstruction account posits that motor simulation serves action monitoring and action prediction. That is, by engaging motor programs leading to the hypothesized outcome of the witnessed action, the observer examines whether their expected perceptual effect matches the observed movement. Thus, mirroring processes should be reliably in place when observers witness instrumental actions directed at realizing particular behavioral goals. However, such action monitoring is not required in the case of referential actions that are performed to direct the observer's attention towards objects and events that could be referents of the ongoing communication. While at the level of inferences, both types of action interpretation return the partial contents of the actor's intentions (i.e., desired end states and to-be-communicated information, respectively), and, at the level of perceptual analysis, both types of actions can be conceived as transitive object-directed movements (e.g., grasping and pointing, respectively), the links between these levels of representation are formed by different mechanisms (i.e., assessment of action efficiency and recognition of communicative intentions, respectively). Accordingly, different neural systems might be involved in processing actions that were given these two types of interpretation, with a stronger involvement of the motor system in response to actions construed as instrumental and requiring motor emulation of hypothesized goals than in response to actions construed as

communicative and requiring the identification of those aspects of the environment that will be target of the communicative exchange. Thus, the directionality of causal connections between motor mirroring and action understanding can be tested by comparing the levels of motor activity across instrumental and communicative actions. If a particular kind of action interpretation is a prerequisite for the engagement of the observer's motor system, then the selected interpretation should directly modulate the level of motor activation.

Observation of transitive goal-directed actions often elicits stronger motor mirroring than observation of intransitive actions (e.g., Muthukumaraswamy et al., 2004). However, the neural mechanisms involved in interpreting transitive referential actions remain largely unexplored (but see Andric et al., 2013; Pierno et al., 2009). Some evidence for the causal role of action interpretation in motor mirroring comes from research exploring motor activation during observation of actions with no clear goals. For example, reaching behind an occluder elicits motor activation in monkeys when the animals know that there is an object placed behind the occluder; yet, the same reaching movements do not result in a mirroring response when the animals know that no object is there (Umiltà et al., 2001). Similarly, human infants display mu suppression while watching a hand reaching behind an occluder, but not when a matching action is performed with the back of the hand (Southgate et al., 2010). Finally, in human adults motor activation is higher for actions with obvious goals relative to actions whose goals are not straightforward to understand (e.g., Chaminade et al., 2001; Johnson-Frey et al., 2003). On the basis of these results, one could argue that it is transitivity (object-directedness), rather than interpretation of goal-directedness, that elicits or facilitates motor mirroring of observed actions. Direct comparison of transitive object-direct actions that invite radically different interpretations, such as grasping, which is instrumental to retrieving an object, and pointing, which aims at sharing some information about an object, would allow us to address this objection.

The current study was designed to investigate whether action interpretation requires the involvement of the motor system by directly contrasting motor activation in response to observing different types of actions that share formal characteristics, such as transitivity and object-directedness, and whose interpretation is easily accessible to human adults. We contrasted these particular actions because they are typically interpreted as instrumental (i.e., grasping, reaching, e.g., Woodward, 1998; Yoon et al., 2008) or referential and communicative (i.e., pointing, e.g., Kita, 2008) since early in ontogeny. Although the exact kinematics (e.g., Marteniuk et al., 1987; Carnahan et al. 1993; Castiello, 2001) and neural bases underlying the production of these actions (Grafton et al., 1996) differ, all are object-directed (transitive) and have comparable trajectory of movement that is available for mirroring to human adults. To assess motor activation, we measured the mu rhythm (also known as the resting state sensorimotor alpha rhythm, 8-12 Hz) through electroencephalography (EEG). Attenuation (or suppression) of this rhythm is typically observed while participants execute movements or observe biological movements executed by others (e.g., Babiloni et al., 1999; Hari et al., 1998; Muthukumaraswamy et al., 2004) and is considered to reflect the activation of the sensorimotor cortex (Arnstein et al., 2011; Hari & Salmelin, 1997). If the recruitment of the sensorimotor system during action processing is conditional on instrumental action interpretation, mu suppression should be observed only in response to grasping and reaching, but not in response to pointing.

However, it should be acknowledged that different levels of mu suppression during observation of grasping, reaching, and pointing actions could in principle result not from differences in action interpretation, but from the fact that the motor system is more strongly engaged in response to actions that require prehension and more precise motor planning (e.g., to adjust the shape of the hand to the particular target object) than in response to simple pointing actions that require little or no adjustment to the environment. Such is the case

during execution of hand actions: more brain areas are recruited while participants conduct a precision grip relative to a power grip (Begliomini et al., 2007; Ehrsson et al., 2000). To ensure that our results are not by-products of the larger neural population involved in processing prehensile actions, we introduced an additional experimental manipulation: participants were presented with speech (*Look!*) or a matched pure tone prior to the onset of the action. Our rationale was that the presence of communicative signals would bias participants to suppress the instrumental interpretation of the observed acts (and/or construe them as referential), even though the same sequences of movements might not elicit this interpretation outside of communicative contexts. Under this assumption, identical grasping actions should be interpreted differently depending on the preceding sound stimuli, thus affecting levels of motor activation. Specifically, if the presence of speech changes the interpretation of grasping from instrumental to referential, less mu suppression should be recorded in the presence of speech than in the presence of a pure tone. These predictions were tested across two experiments that differed solely with respect to how the experimental trials were blocked for presentation: in Experiment 1 trials were blocked by sound and action, while in Experiment 2 trials were blocked only by sound. Thus, while in Experiment 1 the participants could predict what action they were about to see in each trial, in Experiment 2 this prediction was not available to them.

2. Methods

2.1. Participants

Sixteen healthy paid student volunteers participated in Experiment 1. The data from 15 participants were included in the current analyses (mean age: 22 years, range: 18 to 27 years). One participant was excluded from the analysis due to excessive blinking artifacts, which contaminated more than 50% of the trials. Eighteen volunteers participated in

Experiment 2. The final sample consisted of 16 participants (mean age: 22 years, range: 19 to 25 years). One participant was excluded from the analysis due to excessive blinking, and another one due to sweat artifacts, which contaminated more than 50% of the trials. All participants were healthy, right-handed, and had normal or corrected-to-normal vision (as established using verbal reports). All gave written informed consent to participate in the current research prior to the experimental session. The experiments were approved by the local ethical committee.

2.2. Stimuli and Apparatus

Participants sat in a sound-proof electrically shielded EEG booth at a viewing distance of approximately 80 cm from a 24-in. LCD computer screen (60 Hz refresh rate, 1920 x 1080 px resolution), with a response pad in their hands. Two loudspeakers positioned side by side behind the screen were used to deliver auditory stimuli.

Visual stimuli were video clips depicting a human hand performing one of the three object-directed actions: (1) grasping an object, (2) reaching for an object without grasping it, and (3) pointing to an object with the extended index finger (example clips are available in the Supplementary Material). All the clips had the same structure and timing: first, the object was presented in isolation in the center of a dark surface (1150 ms), then the object was targeted by one of the three actions performed by a hand approaching it from the top of the display (1150 ms), and finally the outcome of the action was presented as a still frame (300 ms), for a total trial length of 2600 ms (Figure 1). The hand made physical contact with the object only in the grasping condition, while it stopped before touching the object in the reaching and pointing conditions. The three types of actions were filmed separately. In the reaching condition, the actor was instructed to move her hand as if she was to grasp the object but could not stretch out her arm enough to touch it. In all clips the movement lasted

exactly 1150 ms, thus resulting in a slightly higher velocity of grasping over reaching and pointing actions due to the longer trajectory of movement when the actor had to touch the object.

Fifteen purpose-made similar-sized objects novel to participants were used to record the clips. Clips were 960 x 540 px in size and were positioned in the center of the screen against a grey background, thus subtending approximately 12° of visual angle. A white fixation dot (5 x 5 px) was displayed in the center of the display. In each video clip, the action was preceded by one of the two types of auditory stimuli: a speech or a non-speech sound. The speech stimuli were three tokens of the phrase "Nézd csak!" ("Look!") recorded by a female Hungarian native speaker. The non-speech stimuli were created by concatenating two pure tones matched in frequency and duration with constituent words of the respective speech stimuli (i.e., "nézd", "csak"). The auditory stimuli started 50 ms after the onset of the clip and lasted from 650 to 677 ms. Each experimental trial comprised a baseline stimulus (800 to 1200 ms, randomly determined), action video clip, and a blink stimulus (800 to 1500 ms, randomly determined). The baseline stimuli were rotating screen-saver like animations. The blink stimulus was a "most pislogj" ("blink now") instruction displayed in the center of the screen against a grey background. It was administered to minimize contamination of the EEG signal by blink artifacts.

The video stimuli were created with Apple Final Cut Pro and the sound was edited using Praat (http://www.fon.hum.uva.nl/praat/). The stimuli presentation was administered using MATLAB 2014b with the Psychophysics Toolbox Version 3 (http://psychtoolbox.org).
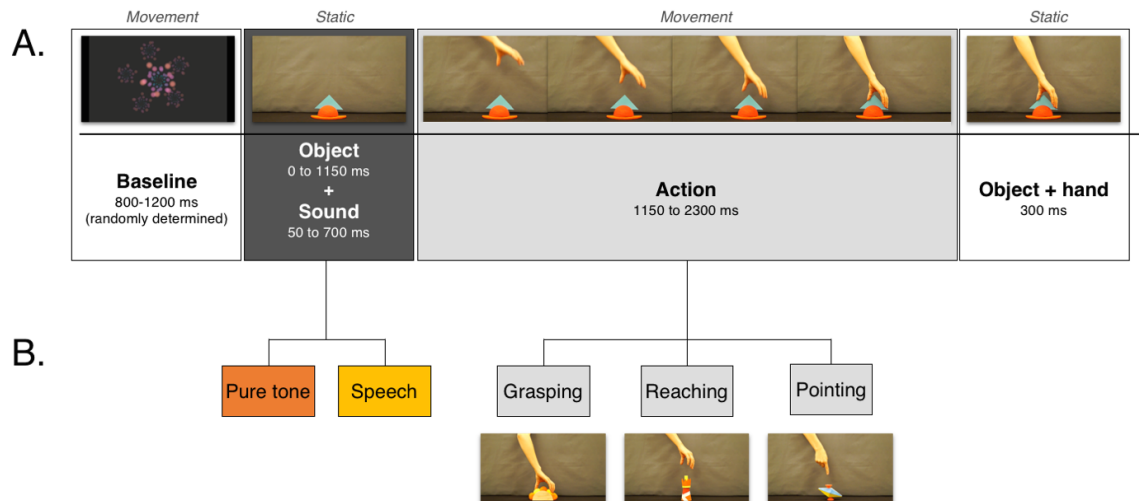
Fig. 1. Experimental design. (A) The time course of the stimuli. Within each trial, the baseline period was followed by a 2600 ms clip depicting a target action. The clips consisted of three phases: object presentation (1150 ms), object-directed action (1150 ms), outcome presentation (300 ms). (B) The experimental conditions and the outcomes of the three target actions. The onset of one of the three object-directed actions (grasping, reaching, or pointing) was preceded by a sound stimulus (speech or pure tone) delivered during the object presentation period.

2.3. Design and Procedure

The study consisted of 6 experimental conditions (2 x 3) defined by combinations of communicative and non-communicative sounds (i.e., speech vs. pure tone) with video clips depicting grasping, reaching, and pointing actions. In Experiment 1, there were 6 types of experimental blocks, in which one type of action was paired with one type of sound (i.e., *pure tone – grasping, pure tone – reaching, pure tone – pointing, speech – grasping, speech – reaching, speech – pointing)*. The sequence of 6 blocks was repeated 3 times (each time in a different order, randomly determined), for a total of 18 blocks. Each block comprised 20 trials (total number of trials: 360). In Experiment 2, trials were blocked solely by sound, so that in each block participants saw 21 trials depicting grasping, reaching, and pointing actions ($n = 7$ trials per action type) intermixed with the constraint that the same action could not be presented more than twice in a row. In total, 18 blocks were administered (total number of trials: 378).

While viewing the object-direct actions, participants performed a cover task to maintain their concentration during the experimental session. Participants were instructed to count how many times a particular object appeared on the screen during the block. In each block, we presented clips featuring only 5 preselected objects (randomized). Before the block started participants were shown the picture of the target object (randomly chosen from the set of 5 objects). After the block ended, they had to enter the response (forced choice between three options: 3, 4, or 5). The objects appeared with different prefixed frequency (in Experiment 1: 2 objects were presented on 3 and 5 trials, respectively, and 3 remaining objects were presented on 4 trials each; in Experiment 2: 1 object was presented on 3 trials, 2 objects were presented on 4 trials each, and 2 remaining objects were presented on 5 trials each). The frequency of appearance of the target object was random with the constraint that it could not be the same on more than two consecutive blocks.

Participants were presented with written and verbal instructions to remain still and maintain their gaze on the centrally displayed fixation point to eliminate artifacts due to movement and saccades. They were also instructed to blink only when the blink stimulus was presented on the screen. Prior to the start of the experimental blocks, participants completed 5 silent training trials.

2.4. EEG Acquisition and Analysis

The continuous EEG was acquired using HydroCel Geodesic Sensor Nets composed of 128 Ag/AgCl sensors (Electrical Geodesics Inc., Eugene, OR, USA). Recordings were referenced to the vertex (Cz in the 10-20 system). The ground electrode was at the rear of the head (between Cz and Pz). The data were low-pass filtered online during the amplification process at 200 Hz and digitized with 500 Hz sampling rate. Impedances were kept below 40 KΩ.

The continuous EEG was segmented from 1200 ms before to 3200 ms after the onset of the visual stimulus (i.e., action clip). Only segments with the voltage range lower than 70 µV within a sliding window of 80 ms and without ocular artifacts (saccades: voltage exceeding 55 µV, eye blinks: voltage exceeding 140 µV, at the horizontal EOG channels) were included in the analyses. Segments were excluded from the analyses if more than 10% of channels were contaminated by artifacts. Participants contributed on average 46 artifact-free trials in Experiment 1 (median: 47, range: 21 to 59) and 48 artifact-free trials in Experiment 2 (median: 49, range: 14 to 63). There were no significant differences in the number of artifact-free trials across conditions.

Following the interpolation of bad channels, the data were re-referenced to average reference. In order to assess mu suppression, we performed time-frequency analyses by applying a continuous wavelet transformation with Morlet wavelets for frequencies in 5-20 Hz range with 1 Hz resolution. Then, the absolute value (i.e., the amplitude, and not the power) of the resulting complex coefficients was calculated (Csibra et al., 2000). Average event-related activation was computed for each participant by taking the mean across the trials for each experimental condition. To remove the distortion induced at the epoch edges by the wavelet convolution, 700 ms of the buffer data were removed at the beginning and at the end of each segment. The activation was baseline corrected by subtracting the average amplitude of the 500 ms prestimulus baseline from the whole epoch at each frequency. The time-frequency analysis was conducted in MATLAB 2014b using a package of custom-made scripts (WTools, see Parise & Csibra, 2013; EEGLAB, v. 13_4_3b). To assess mu suppression, we used two scalp regions located over sensorimotor cortices and selected around the C3 and C4 sites (e.g., Babiloni et al., 2002; Muthukumarswamy et al., 2004; Muthukumarswamy & Johnson, 2004). Each region comprised a cluster of 8 channels (left: 30, 31, 36, 37, 41, 42, 53, 54; right: 79, 80, 86, 87, 93, 103, 104, 105). The average activity in

the alpha range (8-12 Hz) was computed for these two regions in the 1400-2300 ms time window, corresponding to the interval when the action was executed. Although the first frame when the hand appeared on the screen was at 1150 ms after the stimulus onset, we decided to shift the beginning of the time window of analysis due to the distortion in the signal created by the wavelet computation (i.e., around 250 ms at 8 Hz).

3. Results

3.1. Behavioral results

Participants tracked the target objects with high accuracy, providing 90% correct responses in Experiment 1 (SD = 6%) and 91% in Experiment 2 (SD = 6%). In neither of the two experiments was their performance was affected by the type of block (Experiment 1: $F(5,70) = 1.000, p = .424$; Experiment 2: $t(15) = 0.735, p = .474$).

3.2. Electrophysiological results

3.2.1. Experiment 1. Sensorimotor alpha responses are shown on Figure 2A. Mu suppression was analyzed in a three-way repeated-measures ANOVA with hemisphere (left vs. right), sound (speech vs. pure sound), and action (grasping vs. reaching vs. pointing) as within-subject factors. This analysis yielded a significant interaction between sound and action, $F(2,28) = 4.052, p = .046$. Follow-up paired-samples t tests revealed that this interaction was due to the fact that there was a significant difference between pure tone and sound conditions only during the observation of grasping, $t(14) = 2.282, p = .039, d = 0.534$ but not during the observation of reaching, $t(14) = 1.245, p = .233$, or pointing, $t(14) = 1.024$, $p = .323$. A separate series of one-sample t tests found that only the mu suppression recorded during the observation of grasping preceded by pure tone sounds was significantly different from baseline, $t(14) = 2.191, p = .046$.
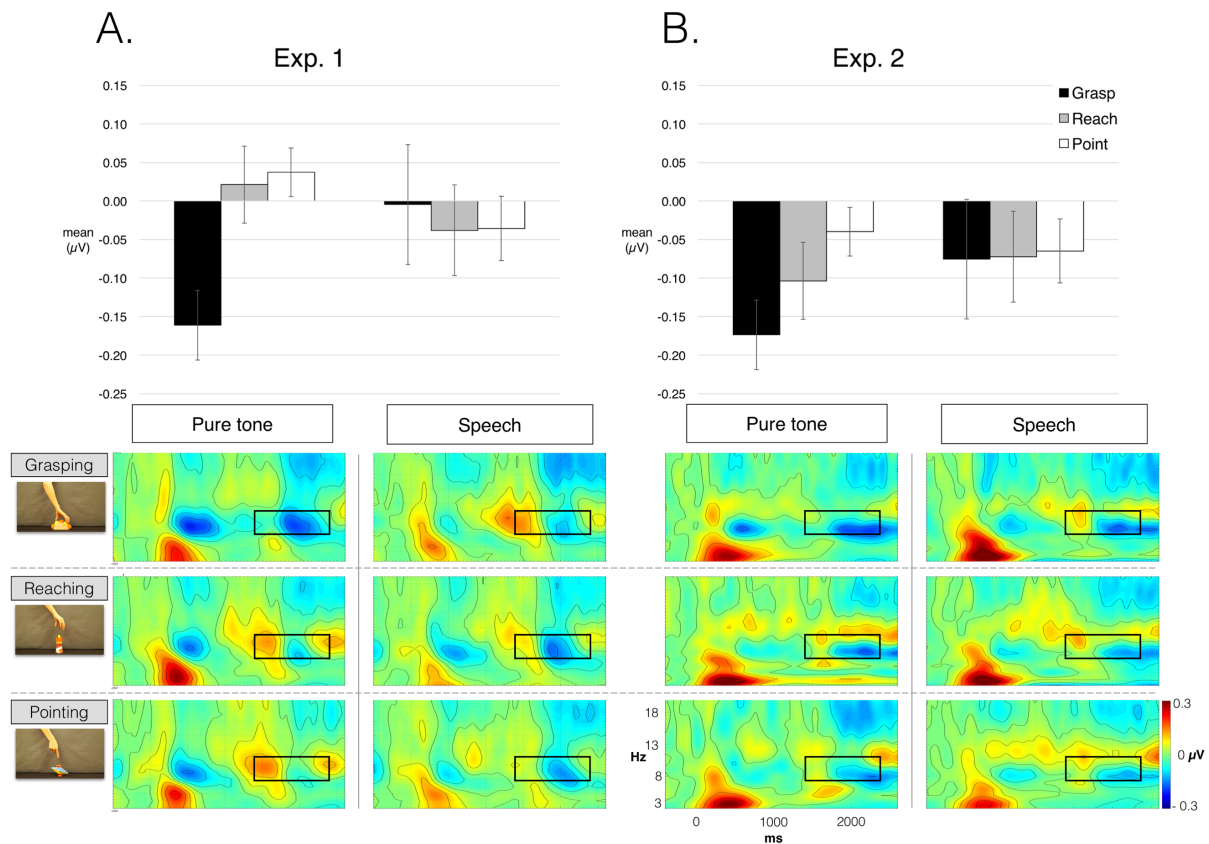
Fig. 2. Summary of the EEG results. (A) Experiment 1, (B) Experiment 2. The time frequency-plots represent the changes in the oscillatory activity (3-20 Hz) during the action observation recorded over C3 and C4 sites. The time-frequency plots show baseline-corrected activity averaged across all participants as a function of the observed action (grasping, reaching, or pointing) and the sound preceding the onset of the action (pure tone or speech). The zero point represents the beginning of the experimental trial. The black squares indicate the time-frequency window used for the analysis of the mu suppression. The bar plots represent the mean amplitude of the oscillatory activity during the window of analysis. Error bars represent SEM.

These results indicate that participants recruited their sensorimotor system (as indexed by mu suppression from baseline to action observation period) only while observing actions that could be unambiguously interpreted as instrumental (i.e., grasping preceded by pure tone sounds). When the context suggested a referential interpretation of the observed action, either due to the semantics of the witnessed gestures (i.e., pointing) or to the presence of speech, there was no sign of significant motor activation. This pattern of results suggests that action interpretation is not dependent on the observer's motor system and that the presence of

subsequent motor activation is conditioned by this interpretation: only conceiving of an action as instrumental, but not as referential, leads to the recruitment of sensorimotor cortices during action observation.

  3.2.2. Experiment 2. Sensorimotor alpha responses are shown on Figure 2B. A repeated-measures ANOVA with hemisphere (left vs. right), sound (speech vs. pure sound), action (grasping vs. reaching vs. pointing) on mu suppression revealed a significant interaction between sound and action, $F(2,30) = 3.746$, $p = .035$. This interaction was driven by the activation during the observation of grasping actions: there was significantly more suppression when grasping was preceded by speech than when it was preceded by pure tones, $t(15) = 2.377$, $p = .031$, $d = 0.36$. The exposure to different types of sound did not lead to significant differences when participants observed reaching, $t(15) = .162$, $p = .874$, or pointing, $t(15) = .519$, $p = .612$. Only when grasping actions followed pure tones, the activation during action observation was significantly different from the activation recorded during baseline period, $t(15) = 2.313$, $p = .035$.

  This pattern of results mirrors that of Experiment 1, thus providing additional evidence that the involvement of sensorimotor cortices during action observation was conditional on the interpretation of the observed action as instrumental. In order to prevent participants from pre-activating action interpretation mechanisms before they were exposed to the action to be interpreted, in Experiment 2 the trials were blocked solely by sound type, and not by both sound and action type as in Experiment 1. The fact that we found the same results suggests that action anticipation is not necessary for the mu suppression response to build up.

4. Discussion

Despite numerous reports of motor mirroring in the observer's brain during the observation of actions performed by others, it remained unclear whether and to what extent this mirroring mechanism contributes to the understanding of observed actions. Two opposite views on its functional significance have emerged in the literature: the mirror-based account of action understanding proposes that the recruitment of the observer's motor system is a precondition of action understanding, while the action reconstruction and the social responding accounts argue that motor mirroring is one of the outcomes of action understanding. Here, we put these competing accounts to the test using an action-observation task, measuring motor mirroring by the modulation of the sensorimotor alpha-band EEG activity (i.e., mu suppression).

Across two experiments, participants displayed the activation of their motor system, as indexed by mu suppression, in response to viewing grasping actions, but not in response to viewing pointing or reaching actions. Besides sharing the same formal characteristics (i.e., transitivity, object-directedness) that were previously shown to elicit higher levels motor mirroring than intransitive actions not involving objects (e.g., Nishitani & Hari, 2000), all presented actions were part of the motor repertoire of human adults and were familiar to the participants. Therefore, the recorded differences in the motor activity could not stem from differences in the structural description of the presented actions (e.g., whether a particular hand movement targets an object or not), availability of corresponding motor representations, or inaccessibility of relevant action interpretations. The fact that motor activation was only observed under circumstances that normally result in a particular interpretation of the observed behaviors (i.e., instrumental rather than referential) indicates that the activation of the motor system follows in time, and is modulated by, the observers' interpretation of ongoing actions. The proposal that action interpretation is achieved through mechanisms that

directly and automatically match observed movements with the observer's own motor representations is inconsistent with the current findings.

The fact that the observation of pointing, a conventionalized communicative act, did not result in increased mu suppression (as compared to baseline) is in line with the idea that the interpretation of communicative acts does not involve the emulation of hypothesized action goals (or end states) in the observer's motor system. Rather, in communicative contexts the observer's cognitive resources are focused on identifying and processing those aspects of the world that are potential referents of the actor's communicative intention, which does not directly engage the motor circuitry.

Furthermore, the motor activation elicited by the observation of grasping actions was modified by the presence of a communicative signal before the onset of the action. Specifically, mu suppression was observed in response to grasping actions that were preceded by pure tones, but not in response to identical grasping actions were preceded by participant-directed speech (i.e., *Look!*). This finding suggests that the ostensive nature of the auditory speech signal changed the interpretation of the observed grasping action: the instrumental interpretation that is normally triggered by this action was withdrawn or suppressed. It is likely that this withdrawal of instrumental interpretation was due to the generation of a different interpretation: the communicative context induced the expectation for referential signals rather than for an instrumental action (Csibra, 2010). Crucially, grasping an object could indeed be the first step of picking up and showing the object to the observer, in which case the object would play the role of the referent of a communicative exchange rather than the target of an instrumental action (e.g., Carpenter et al., 1998). Similarly, the fact that motor activation was not elicited by the observation of reaching actions indicates that reaching in the current context was not construed as goal-directed. Although we did not predict this finding, it can be explained given the pragmatics of the current experimental situation: the

fact that the hand would repeatedly stop short of the object could have made the action uninterpretable in terms of its underlying goals.

At first, our conclusion that motor activation is not elicited, or at least that it is attenuated, in response to the observation of communicative actions may be seen as contradicting an earlier finding that reported different modulation of alpha-band oscillatory activation when the actor faced the participants compared to when he was viewed from behind (Kilner, Marchant, & Frith, 2006). An action in front view with implied eye contact could indeed indicate potential social interaction via non-verbal communication. However, two aspects of the results reported by Kilner et al. (2006) suggest that the reported effect was not related to motor mirroring. First, the modulation of alpha-band oscillatory activity was identified over the posterior parietal cortex, not over sensorimotor areas where mu suppression is usually found during execution or observation of actions (Babiloni et al., 2002). Second, the magnitude of alpha-attenuation was not different between 'facing' and 'away' action; rather, action direction (left vs. right) modulated the hemispheric activity differently only in the 'facing' condition. This pattern of results suggests that the orientation of the observed actor had an effect not on the motor system of the observers but on their spatial attention, and Kilner et al. (2006) arrived at a similar conclusion.

Since early infancy, humans are able to determine when they are addressed by communication (Grossman et al., 2008; Senju & Csibra, 2008). This achievement is partly due to sensitivity to a set of specialized signals, such as speech, eye contact and contingent responsiveness to others' actions (Csibra, 2010), that unambiguously indicate that subsequent actions are intended to be communicative (i.e., ostensive signals, after Sperber & Wilson, 1986). Given that human communication employs a variety of channels and means, and that most object-direct actions can be performed with a communicative aim (e.g., tool-use demonstration instead of instrumental tool use), addressees should be ready to extend

referential action interpretation to any behavior that is accompanied by ostensive signals. In the current study, the withdrawal of instrumental interpretation of observed grasping actions was triggered by the presence of speech, but other ostensive signals should be functionally equivalent to speech in modulating action interpretation.

Our data provides evidence that processing of actions that are not interpreted as goal-directed might not induce motor stimulation. However, the idea that the interpretation of instrumental actions might be also causally independent from motor resonance has been recently corroborated by findings from patients with limb dysplasia, a condition wherein upper limbs are absent or severely shortened. Vannuscorps and Camarazza (2016) demonstrated that these individuals could efficiently recognize and anticipate various upper limb actions despite the lack of motor representations and motor programs necessary to simulate the limb movements required to perform them. Furthermore, while watching hand actions aplasic patients born without hands displayed motor activation in the feet and/or mouth areas that corresponded to the effectors they would use to achieve the same goals (e.g., to grasp a container, Gazzola et al., 2007). Thus, motor activation here was the result of, rather than a contributor to, goal identification.

What neural mechanisms underlie action interpretation? The ability to interpret behaviors of others is supported by computations that take place outside of the motor system. For example, traditional cognitive models of action understanding propose that action interpretation is generated in two phases: observed movements are first analyzed in the observer's visual system, and then fed directly into a praxis conceptual system that assigns meaning to them (Rothi et al., 1991). It has been also suggested that action processing might rely exclusively on visual processing implemented in hierarchical feedforward architectures (Giese & Poggio, 2003). However, our findings cannot be fully explained within a strictly feedforward framework. The fact that the interpretation of grasping actions differed

depending on the presence of speech implies that not only the characteristics of the visual stimuli, but also the context of the action execution contributed to its interpretation. This implies that human action interpretation abilities cannot be based exclusively on stimulus-driven processes, and that action cognition relies to some extent on expectations derived from long-term conceptual knowledge (for convergent evidence suggesting that contextual information facilitates action recognition, see Wurm & Schubotz, 2012, 2016).

Although the question of neural substrates of such top-down influences remains open, recent studies have provided evidence supporting claims about abstract action analysis outside of the human motor system. Wurm and Lingau (2015) demonstrated that abstract action representations (e.g., OPEN) that generalize across different object categories (e.g., open a bottle, open a box) and across different perceptually distinct objects within a given category (e.g., open this small bottle, open that large bottle, etc.) are instantiated in parietal and occipitotemporal cortices. In contrast, premotor cortices represent low-level features idiosyncratic to specific instantiations of an observed action.

One important question to be addressed in the light of the current results is what cognitive consequences follow from adopting a particular interpretative stance towards observed behaviors. In particular, to what extent are different types of action interpretations influenced by object representation and conceptualization? Regarding object representation, studies suggest that different information about objects is stored in the working memory depending on whether they were targeted by an action interpreted as instrumental or referential (Marno et al., 2014; Yoon et al., 2008). For example, preverbal infants are more likely to remember the location of a toy in an instrumental context, while in a referential context they preferentially store the information about its appearance (Yoon et al., 2008). Marno and colleagues (2014) demonstrated that such representational biases continue to shape working memory object representations in adulthood. It is less clear, however, whether

action interpretation also shapes the way in which objects are conceptualized. For example, it is possible that objects picked out through referential actions are construed as referents of the subsequent communication, while no such interpretation is attributed to objects selected through instrumental actions. Such interpretative strategy could considerably constrain infant word learning: infants could more easily learn new words for objects introduced in referential rather than in instrumental contexts (Pomiechowska & Csibra, 2016).

In sum, we have provided evidence that motor activation in the observer's brain is present when witnessed movements are construed as instrumental, and that such activation is absent when they are conceived as communicative, implicating that the involvement of motor system in action processing is not a prerequisite but a result of action interpretation. The two action interpretation systems that represent object-directed actions as instrumental and referential, respectively, are operational already in early childhood, support action interpretation in adulthood, and are likely to be implemented in distinct neural mechanisms. Our findings challenge the widely received view that motor resonance, or simulation of the observed movements in the observer's own motor system, is crucial to action understanding. The fact that action interpretation flexibly changes depending on the presence of ostensive communicative signals implies that action processing mechanisms make use of both bottom-up and top-down mechanisms. The neural architecture of such action processing system should be elucidated by further research.

Acknowledgements

References

Andric, M., Solodkin, A., Buccino, G., Goldin-Meadow, S., Rizzolatti, G., & Small, S. L. (2013). Brain function overlaps when people observe emblems, speech, and grasping. *Neuropsychologia*, 51, 1619-1629.

Arnstein, D., Cui, F., Keysers, C., Maurits, N. M., & Gazzola, V. (2011). µ-suppression during action observation and execution correlates with BOLD in dorsal premotor, inferior parietal, and SI cortices. *The Journal of Neuroscience*, *31*, 14243-14249.

Babiloni, C., Babiloni, F., Carducci, F., Cincotti, F., Cocozza, G., Del Percio, C., ... & Rossini, P. M. (2002). Human cortical electroencephalography (EEG) rhythms during the observation of simple aimless movements: a high-resolution EEG study. *Neuroimage*, *17*, 559-572.

Babiloni, C., Carducci, F., Cincotti, F., Rossini, P. M., Neuper, C., Pfurtscheller, G., & Babiloni, F. (1999). Human movement-related potentials vs desynchronization of EEG alpha rhythm: a high-resolution EEG study. *Neuroimage*, *10*, 658-665.

Begliomini, C., Wall, M. B., Smith, A. T., & Castiello, U. (2007). Differential cortical activity for precision and whole-hand visually guided grasping in humans. *European Journal of Neuroscience*, 25, 1245-1252.

Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H. J., & Rizzolatti, G. (2004). Neural circuits underlying imitation learning of hand actions: an event-related fMRI study. *Neuron*, *42*, 323-334.

Carnahan, H., Goodale, M. A., & Marteniuk, R. G. (1993). Grasping versus pointing and the differential use of visual feedback. *Human Movement Science*, 12, 219-234.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the society for research in child development*, i-174.

Castiello, U. (2001). The effects of abrupt onset of 2-D and 3-D distractors on prehension movements. *Attention, Perception, & Psychophysics*, 63, 1014-1025.

Chaminade, T., Meary, D., Orliaguet, J. P., & Decety, J. (2001). Is perceptual anticipation a motor simulation? A PET study. *NeuroReport*, *12*, 3669-3674.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, *25*, 141-168.

Csibra, G. (2008). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition*, *107*, 705-717.

Csibra, G. (2007). Action mirroring and action interpretation: An alternative account. In: P. Haggard, Y. Rosetti, & M. Kawato (Eds.), *Sensorimotor Foundations of Higher*

Cognition. Attention and Performance XXII (pp. 435-459). Oxford: Oxford University Press.

Csibra, G. (2003). Teleological and referential understanding of action in infancy. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *358*, 447-458.

Csibra, G., & Gergely, G. (2007). 'Obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans. *Acta psychologica*, *124*, 60-78.

Csibra, G., & Volein, A. (2008). Infants can infer the presence of hidden objects from referential gaze information. *British Journal of Developmental Psychology*, *26*, 1-11.

Csibra, G., Bıró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science*, *27*, 111-133.

Csibra, G., Davis, G., Spratling, M. W., & Johnson, M. H. (2000). Gamma oscillations and object processing in the infant brain. *Science*, *290*, 1582-1585.

Csibra, G., Gergely, G., Bíró, S., Koos, O., & Brockbank, M. (1999). Goal attribution without agency cues: the perception of 'pure reason'in infancy. *Cognition*, *72*, 237-267.

Daum, M. M., Prinz, W., & Aschersleben, G. (2008). Encoding the goal of an object-directed but uncompleted reaching action in 6-and 9-month-old infants. *Developmental Science*, *11*, 607-619.

Dennett, D. C. 1987 *The intentional stance*. Cambridge, MA: MIT Press.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental brain research*, *91*, 176-180.

Elsner, B. (2007). Infants' imitation of goal-directed actions: The role of movements and action effects. *Acta psychologica*, *124*, 44-59.

Ehrsson, H. H., Fagergren, A., Jonsson, T., Westling, G., Johansson, R. S., & Forssberg, H. (2000). Cortical activity in precision-versus power-grip tasks: an fMRI study. *Journal of neurophysiology*, *83*, 528-536.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in cognitive sciences*, *8*, 396-403.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593-609.

Gazzola, V., van der Worp, H., Mulder, T., Wicker, B., Rizzolatti, G., & Keysers, C. (2007). Aplasics born without hands mirror the goal of hand actions with their feet. *Current biology*, *17*, 1235-1240.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naıve theory of rational action. *Trends in cognitive sciences*, *7*, 287-292.

Grafton, S. T., Arbib, M. A., Fadiga, L., & Rizzolatti, G. (1996). Localization of grasp representations in humans by positron emission tomography. *Experimental brain research*, 112, 103-111.

Jacob, P., & Gergely, G. (2012). Reasoning about instrumental and communicative agency in human infancy. *Rational constructivism in cognitive development*, 59-94.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*, 165-193.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, *4*, 179-192.

Gliga, T., & Csibra, G. (2009). One-year-old infants appreciate the referential nature of deictic gestures and words. *Psychological Science*, *20*, 347-353.

Gredebäck, G., Fikke, L., & Melinder, A. (2010). The development of joint visual attention: a longitudinal study of gaze following during interactions with mothers and strangers. *Developmental science*, *13*, 839-848.

Grèzes, J., Armony, J. L., Rowe, J., & Passingham, R. E. (2003). Activations related to "mirror" and "canonical" neurones in the human brain: an fMRI study. *Neuroimage*, *18*, 928-937.

Grossmann, T., Johnson, M. H., Lloyd-Fox, S., Blasi, A., Deligianni, F., Elwell, C., & Csibra, G. (2008). Early cortical specialization for face-to-face communication in human infants. *Proceedings of the Royal Society of London B: Biological Sciences*, *275*, 2803-2811.

Hari, R., & Salmelin, R. (1997). Human cortical oscillations: a neuromagnetic view through the skull. *Trends in neurosciences*, *20*, 44-49.

Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S., & Rizzolatti, G. (1998). Activation of human primary motor cortex during action observation: a neuromagnetic study. *Proceedings of the National Academy of Sciences*, *95*, 15061-15065.

Hickok, G. (2013). Do mirror neurons subserve action understanding?. *Neuroscience letters*, *540*, 56-58.

Jacob, P. (2008). The scope and limits of enactive approaches to visual experience. *Perception*, *37*, 446-461.

Johnson-Frey, S. H., Maloof, F. R., Newman-Norlund, R., Farrer, C., Inati, S., & Grafton, S. T. (2003). Actions or hand-object interactions? Human inferior frontal cortex and action observation. *Neuron*, *39*, 1053-1058.

Keller, G. B., & Hahnloser, R. H. (2009). Neural processing of auditory feedback during vocal practice in a songbird. *Nature*, *457*, 187-190.

Kilner, J. M., Marchant, J. L., & Frith, C. D. (2006). Modulation of the mirror system by social relevance. *Social cognitive and affective neuroscience*, 1, 143-148.

Kita, S. (Ed.). (2008). *Pointing: Where language, culture, and cognition meet*. Psychology Press.

Marno, H., Davelaar, E. J., & Csibra, G. (2014). Nonverbal communicative signals modulate attention to object properties. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 752.

Marteniuk, R. G., MacKenzie, C. L., Jeannerod, M., Athenes, S., & Dugas, C. (1987). Constraints on human arm movement trajectories. *Canadian Journal of Psychology/ Revue canadienne de psychologie*, 41, 365.

Muthukumaraswamy, S. D., & Johnson, B. W. (2004). Changes in rolandic mu rhythm during observation of a precision grip. *Psychophysiology*, *41*, 152-156.

Muthukumaraswamy, S. D., Johnson, B. W., & McNair, N. A. (2004). Mu rhythm modulation during observation of an object-directed grasp. *Cognitive Brain Research*, *19*, 195-201.

Nishitani, N., & Hari, R. (2000). Temporal dynamics of cortical representation for action. *Proceedings of the National Academy of Sciences*, *97*, 913-918.

Parise, E., & Csibra, G. (2013). Neural responses to multimodal ostensive signals in 5-month-old infants. *PloS one*, *8*, e72360.

Pierno, A. C., Tubaldi, F., Turella, L., Grossi, P., Barachino, L., Gallo, P., & Castiello, U. (2009). Neurofunctional modulation of brain regions by the observation of pointing and grasping actions. *Cerebral Cortex*, *19*, 367-374.

Pomiechowska, B. & Csibra, G. (2016). Referential understanding of pointing actions and its consequences for object representation. International Conference on Infant Studies, New Orleans, LA, US.

Prather, J. F., Peters, S., Nowicki, S., & Mooney, R. (2008). Precise auditory–vocal mirroring in neurons for learned vocal communication. *Nature*, *451*, 305-310.

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

Sinigaglia, C., & Rizzolatti, G. (2011). Through the looking glass: Self and others. *Consciousness and cognition*, *20*, 64-74.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, *2*, 661-670.

Gonzalez Rothi, L. J., Ochipa, C., & Heilman, K. M. (1991). A cognitive neuropsychological model of limb praxis. *Cognitive Neuropsychology*, *8*, 443-458.

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, *18*, 668-671.

Skerry, A. E., Carey, S. E., & Spelke, E. S. (2013). First-person action experience reveals sensitivity to action efficiency in prereaching infants. *Proceedings of the National Academy of Sciences*, *110*, 18728-18733.

Southgate, V., & Csibra, G. (2009). Inferring the outcome of an ongoing novel action at 13 months. *Developmental psychology*, *45*, 1794.

Southgate, V., Johnson, M. H., El Karoui, I., & Csibra, G. (2010). Motor system activation reveals infants' on-line prediction of others' goals. *Psychological Science*.

Southgate, V., Johnson, M. H., Osborne, T., & Csibra, G. (2009). Predictive motor activation during action observation in human infants. *Biology Letters*, rsbl20090474.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Second Edition. Oxford: Blackwell.

Umilta, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., & Rizzolatti, G. (2001). I know what you are doing: A neurophysiological study. *Neuron*, *31*, 155-165.

Vannuscorps, G., & Caramazza, A. (2016). Typical action perception and interpretation without motor simulation. *Proceedings of the National Academy of Sciences*, *113*, 86-91.

Wagner, L., & Carey, S. (2005). 12-month-old infants represent probable endings of motion events. *Infancy*, *7*, 73-83.

Woodward, A. L. (2009). Infants' grasp of others' intentions. *Current directions in psychological science*, *18*, 53-57.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*, 1-34.

Wurm, M. F., & Lingnau, A. (2015). Decoding actions at different levels of abstraction. *The Journal of Neuroscience*, *35*, 7727-7735.

Wurm, M. F., & Schubotz, R. I. (2016). What's she doing in the kitchen? Context helps when actions are hard to recognize. *Psychonomic Bulletin & Review*, 1-7.

Wurm, M. F., & Schubotz, R. I. (2012). Squeezing lemons in the bathroom: contextual information modulates action recognition. *Neuroimage*, *59*, 1551-1559.

Yoon, J. M., Johnson, M. H., & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences*, *105*, 13690-13695.