Title page

Title: Typical predictive eye movements during action observation without effector-specific motor simulation

Running head: Motor simulation in predictive eye movement

Authors: Gilles Vannuscorps [1,2] and Alfonso Caramazza [1,2]

[1] Department of Psychology, Harvard University, Cambridge, MA, 02138, USA

[2] Center for Mind/Brain Sciences, Università degli Studi di Trento, Mattarello, 38122, Italy

Correspondence: Gilles Vannuscorps, Department of Psychology, Harvard University, Williams James Hall, 9th floor, 33 Kirkland St., Cambridge, MA, 02138, USA. Phone: (617) 496-6374. E-mail: gvannuscorps@fas.harvard.edu.

Word count (main text, appendixes, footnotes, and references): 3993

Abstract

When watching someone reaching to grasp an object, we typically gaze at the object before the agent's hand reaches it, that is, we make a "predictive eye movement" to the object. The received explanation is that predictive eye movements rely on a direct matching process by which the observed action is mapped onto the motor representation of the same body movements in the observer's brain. In this paper, we report evidence that calls for a re-examination of this account. We recorded the eye movements of an individual born without arms (D.C.) when watching an actor reaching for one of two different-sized objects with a power grasp, a precision grasp or a closed fist. D.C. showed typical predictive eye movements modulated by the actor's hand shape. This finding constitutes existence proof that predictive eye movements during action observation can rely on visual and inferential processes unaided by effector-specific motor simulation.

Introduction

When watching someone grasping an object, we typically gaze at the object before the agent's hand reaches it, i.e., we make a "predictive eye movement" to the object. What kind of representations and processes underlie these predictive eye movements?

Since the seminal work by Flanagan and Johansson (2003), the received explanation is that predictive eye movements rely on a direct matching process by which the observed action is mapped onto the motor representation of the same body movements in the observer's brain. On this account, this effector-specific motor simulation of the observed action would automatically activate, in the observer's brain, the predictive eye motor programs that accompany the execution of that action (Ambrosini, Costantini, & Sinigaglia, 2011; Ambrosini et al., 2013; Ambrosini, Sinigaglia, & Costantini, 2012; Costantini, Ambrosini & Sinigaglia, 2012a, b; Elsner, D'Ausilio, Gredeback, Falck-Ytter, & Fadiga, 2013; Falk-Ytter, Gredebäck, & von Hofsten, 2006). Two main sets of data are cited in support of this hypothesis. One is the results showing that predictive eye movements during the observation of hand actions are influenced by the state of the motor system: they are delayed by a concurrent finger tapping task (Cannon & Woodward, 2008), modulated by participants' hand position (Ambrosini et al., 2012), and hampered by the transient disruption of parts of the hand action production neural network (Costantini, Ambrosini, Cardellicchio, & Sinigaglia, 2013; Elsner et al., 2013). The other is the finding of a relationship between infants' motor abilities and predictive eye movements (Ambrosini et al., 2013; Falck-Ytter et al., 2006).

However, these observations can also be accommodated by an alternative hypothesis according to which predictive eye movements are mainly driven by visual and inferential mechanisms but can be *modulated* by the state of the motor system (Falck-Ytter, 2012; Eshuis, Coventry, & Vulchanova, 2009). There is indeed some evidence that motor activity can modulate the activity of visual brain areas thought to play a role in the perception of body shape and body movements (e.g., Astafiev et al., 2004). Thus interfering with the motor system could have functional effects upon predictive eye movements because it disrupts normal recursive interactions between the motor and visual brain areas serving other purposes such as action imitation (e.g., Buccino et al., 2004) or short term memory (Moreau, 2013; Vannuscorps & Caramazza, 2016). On this hypothesis, the relationship between infants' motor abilities and predictive eye movements finds a natural explanation on the assumption that motor experience contributes not only to the learning of motor representations but also to the development of spatiotemporal knowledge about the performed action, and that the latter, rather than the former, contributes to action anticipation (Southgate, 2013).

In this study, we used a novel approach to test whether predictive eye movements during action observation *require* a direct matching process. We reasoned that if eye movements during action observation *require* direct matching and effector-specific motor simulation, then, predictive eye movements should be specific to actions that an observer can simulate within his/her own motor system. To test this prediction, we recorded the eye movements of an individual (D.C.) born without upper limbs (bilateral upper limb dysplasia) and no history of upper limb prosthetics or phantom limb sensations when watching an actor reaching for or grasping one of two target objects with a precision or a power grasp.

D.C.'s congenital lack of upper limbs has prevented him from learning the motor programs associated with the execution of the typical reaching and grasping actions shown in this study. Thus, unless one assumed that motor programs of actions such as "grasping with the thumb and the index finger" were innate rather than learned by trial and error during development, we can conclude that DC does not have upper limb motor representations on which the upper limb movements shown in this study could be mapped. Therefore, if predictive eye movements during action observation rely on direct matching and effector-specific motor simulation, D.C.'s eye movements when observing hand actions should be reactive, not predictive. If, however, predictive eye movements rely on visual and inferential mechanisms, then, D.C.'s eye movements should be analogous to those of typically developed participants.

To ensure sensitivity in detecting possibly abnormal predictive eye movements in D.C, we measured participant's eye movements when observing an actor reaching for one of two potential target objects (a small or a large one) with either a closed fist or a hand shape appropriate for grasping one of them (a precision or a power grasp). This experimental design provided three main advantages over designs involving only one target object: (1) participants did not know the intention of the agent (which target and which hand shape) in advance, thereby ensuring that predictive eye movements, if any, were driven by the processing of cues inherent to the stimulus; (2) it allowed testing whether participant's eye movements are modulated by the presence of the actor's hand shape information, a modulation considered to be evidence for the crucial role of motor simulation in action perception (Ambrosini et al., 2011; Costantini et al., 2013; Falck-Ytter, 2012); (3) this design has been previously shown to be very sensitive in detecting even subtle impairments of predictive eye movements (Ambrosini et al., 2012; Costantini et al., 2013).

Material and Methods

Participant

D.C. is a 53 year-old man whose performance in various experiments has already been reported (Vannuscorps, Pillon & Andres, 2012; Vannuscorps, Andres & Pillon, 2013, 2014). He has a master's degree in psychopedagogy. He has a congenital bilateral upper limb dysplasia due to in-utero thalidomide exposure. Congenital abnormalities include aplasia of the left upper limb (i.e., the most severe form of dysplasia characterized by a complete absence of arm, forearm, hand and fingers) and, on the right side, a shortened right arm (± 12 cm humerus or ulna) directly fusionned to a hand composed of fingers 1 and 3 (shoulder, elbow and wrist joints absent or not functional). D.C. can move his right upper extremity as a whole by a couple of centimeters in every direction and hold light objects by squeezing them between his chest and shortened limb. However, the absence of most of the typical bones and joints in D.C.'s right upper extremity has prevented typical muscle insertion and development. Therefore, his finger mobility is too limited to allow him to make a precision or palm grip and he cannot grasp or manipulate any object. Instead, DC developed fine motor skills of the feet from early childhood which allow him to use his feet for many typically hand related actions of daily life (e.g., writing with a pen, typewriting, eating with a fork, washing himself, etc.). Physical and mental development was otherwise normal. D.C. never wore prosthetics and reports no history of phantom limb sensations or movements. D.C. has impaired vision of the left eye (visual acuity = 0.5/10) due to acquired eye injury but normal vision of the right eye. Ten normally developed control subjects also participated in the study (3 women, participants S1 and S6 were left-handed, mean age = 30). The study was approved by the biomedical ethic committee of the Cliniques Universitaires Saint-Luc, Brussels, and all participants gave their written informed consent prior to the study.

Procedure and stimuli

Stimuli, provided by Marcello Costantini (Ambrosini et al., 2011; Costantini et al., 2013), were video-clips showing the right hand of an actor, viewed in a third person perspective from the left side, moving forward along ~70 cm to reach one of two potential objects (a small or a large target object) positioned on a table 10 cm apart from each other with either a closed fist ("NO-SHAPE" movement condition) or a hand shape ("SHAPE" movement condition) appropriate for grasping them (a precision or a power grasp, respectively). Four different object layouts were used to counterbalance the hand trajectories (see Ambrosini et al., 2011). The total number of different videos was thus 16 (2 targets x 2 movement condition x 4 layouts). All videos had the same timing: they started with a 1000 msec fixation cross superimposed on the actor's hand (fixation phase), then the actor's hand moved toward one of the two target objects for 1000 msec, then the last frame of the video was shown for 500 msec.

During the experiment, participants were seated with the head stabilized by a head and chin rest at 52 cm from a 17 inches computer screen (395 x 290 mm, 1280*1024 pixels, 60 Hz). In that way, all the videos were sized 21.4° x 16.16° of visual angle (197.5 x 136 mm). Every trial started with the presentation of a central white fixation cross lasting 500, 1000 or 1500 msec (balanced across conditions) that the participants were asked to fixate. Then, the video was displayed and participants were asked to fixate the white cross superimposed on the actor's hand (fixation phase) and then to "simply watch the video" (Ambrosini et al., 2011). Participants performed 2 blocks of 48 trials preceded by a familiarization phase and 10 practice trials. Both blocks were preceded by standard 9-point calibration and validation of the participant's eye-gaze position. In each block, the four layouts, two movement conditions and two target objects, were mixed in a different random order.

The presentation of the stimuli was controlled by E-Prime software (Psychological Software, 2002, Pittsburgh, PA) and participants' eye movements were recorded by an Eyelink 1000 desktop mounted eye tracker (SR Research, Canada; sampling rate of 1000 Hz, average accuracy range 0.25-0.5°, gaze tracking range of 32° horizontally and 25° vertically).

Analyses and results

During the recording session, gaze traces were parsed by means of a saccadic velocity-threshold identification algorithm implemented in Eyelink 1000. EyeLink® Data Viewer (SR

Research, Canada) was used to extract the timing and position of participants' fixations in each trial. In order to compensate for the drift in the eye tracker signal offset over time, a corrective adjustment was applied to the position of participants' fixations based on the difference between the raw position of the participant's eye at the beginning of each trial and the position of the fixation cross they were asked to fixate (Hornof & Halverson, 2002).

Regions of interest (ROI) centered on the actor's hand (hand ROI) and on the target object (target ROI) were created for the 16 different video-clips. The ROIs were circles 0.2° larger than the object to compensate for noise in eye tracking (Costantini et al., 2012a; 2013). Then, for each trial we calculated the time of the first saccade from the hand ROI (gaze-onset time) and that of the first fixation on the target ROI (gaze-arrival time) relative to the hand-movement onset and offset (Ambrosini et al., 2011).

These analyses were performed over two dependent variables: an "accuracy" measure corresponding to the percentage of trials in which participants' gaze fixated on the target ROI before the hand movement offset (Ambrosini et al., 2011) and a "latency" measure computed for each participant by subtracting their mean gaze arrival time from the hand movement offset time (thus, the larger the latency, the larger the anticipation). Trials in which participants did not fixate in the hand ROI at the hand movement onset (end of the fixation phase) were discarded from the accuracy analyses (2% and 5% of the trials in DC and controls, respectively). For the latency analysis, we further discarded trials in which participants did not fixate the target ROI before the end of the trial (3% and 38% of the trials in DC and controls, respectively) and trials in which the gaze arrival time deviated from the participant's mean by 2 SD in each movement condition (3.5% and 3.9% of the trials in DC and controls, respectively).

Distinct analyses were performed to test whether D.C.'s eye movements when observing hand actions were "predictive" (i.e., with the whole set of stimuli) and to test whether they were influenced by the availability of information about the actor's hand shape (i.e., we contrasted the stimuli with and without hand shape information).

General analyses of D.C.'s and controls' eye movements

The direct matching hypothesis predicts that D.C.'s eye movements when observing hand actions should be reactive, not predictive. Contrary to this expectation, we found that D.C's eye movements were predictive: he fixated on the target ROI before the hand offset in 82% of trials and, on average, fixated the target ROI 291 msec before the hand offset (Figure 1). D.C. had in fact a non-significant tendency (both modified t tests, ts (9) < 2, ps > 0.05; Crawford & Howell, 1998) to fixate on the target ROI before the hand offset more often (82% of the trials) and earlier (mean anticipation: 291 msec) than the controls (48% and 179 msec). To further investigate this result, we then used a one-sided Bayesian t test (with a default Cauchy prior width of $r = 0.707$ for effect size on the alternative hypothesis, as specified by Rouder, Morey, Speckman, & Province, 2012) with items as the random variable to quantify the evidence in the data for the hypotheses that D.C. anticipates the outcome of the agent's action (1) less accurately or (2) less rapidly than the controls (H1), and for their corresponding null hypotheses (H0). These analyses, performed with Morey and Rouder's (2015) BayesFactor R package, yielded two Bayes factors (H1/H0) of 0.06, indicating that the observed data were 15 times more likely under the null hypothesis than under the alternative hypothesis. According to Jeffreys (1961), this constitutes *strong evidence* in favor of the null hypothesis.

Effect of the availability of information about the actor's hand shape

The effect of hand shape information on participants' eye movements is considered to be evidence for the crucial role of motor simulation in action perception. Our results, shown in

Figure 2, call for a reexamination of this claim: D.C. was more accurate (SHAPE = 93.62%, NO-SHAPE = 70.21%, chi^2 (1) = 8.69, p = 0.003) and made earlier fixations on the target (SHAPE = - 344 msec, NO-SHAPE = -231msec) when information about the actor's hand shape was available, and the influence of the presence of the actor's hand shape was larger in D.C. than in the control group in both accuracy and latency (Bayesian Standardized Difference Test, both *p*s < 0.05; Crawford & Garthwaite, 2007).

We further investigated this difference between D.C. and the control participants. A reliably larger influence of the presence of the actor's hand shape information in D.C. would indicate that engaging in motor simulation is not only not necessary but could interfere with predictive eye movements. To that end, we first conducted Bayesian ANOVAs on participants' accuracy and latency with items as the random variable to quantify the evidence in the data for the hypotheses that there is an interaction between group ("D.C." vs "CONTROLS") and stimuli ("SHAPE" vs "NO-SHAPE") and the corresponding null hypotheses according to which there is no interaction (Rouder et al., 2012). The comparison of the Bayes factors of the ANOVA models including only the main effects (H0) and the ANOVA models including also their interaction (H1, with a default Cauchy prior width of r = 0.707 for effect size) performed with the BayesFactor R package (Morey & Rouder, 2015) provided only anecdotal support for the interaction model in the latency analysis (1:1.4) and for the absence of interaction in the accuracy analysis (1.5:1).

Then, we also looked at the data of an independent group of 15 typically developed healthy participants reported by Ambrosini et al. (2011, see Table 1). This group of participants tended to be more influenced by the presence of an actor's hand shape that the control participants tested in this study (see Table 1). As these participants were tested with the same stimuli and procedures as the participants tested in our study, this difference is likely the result of the large inter-individual differences in predictive eye movements that was already visible among our control participants (Figures 1 and 2). Importantly, however, the influence of the presence of the actor's hand shape in this independent group of 15 typically developed participants was very similar to that found in D.C. (see Table 1). These two additional analyses suggest that D.C.'s predictive eye movements were very similar to those of control participants.

Table 1

Accuracy and mean anticipation time for the whole set of stimuli, for the stimuli of the NO-SHAPE and SHAPE movement conditions and their difference in DC, the control participants and 15 typically developed healthy participants reported by Ambrosini et al. (2011).

		D.C.	Control participants	Control participants from Ambrosini et al. (2011)
Accuracy	All stimuli	82 %	48 %	73.2%
	Shape	93.6 %	51.75 %	82%
	No-Shape	70.2 %	44.19 %	65%
	Difference	23.4 %	7.56 %	17%
Anticipation	All stimuli	- 291 msec	-179 msec	n.a
	Shape	- 344 msec	-186 msec	-143 msec
	No-Shape	- 231 msec	-171 msec	+10 msec
	Difference	113 msec	15 msec	153 msec

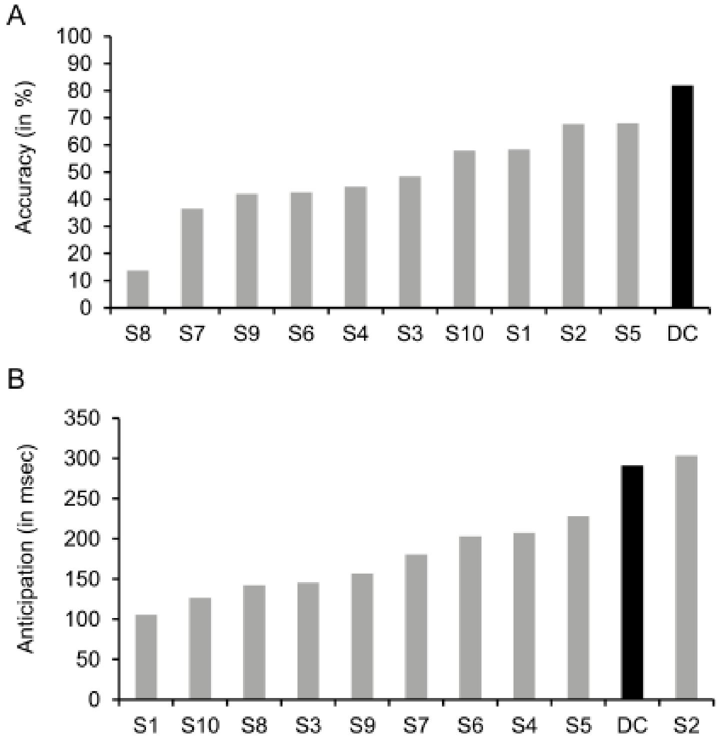

Figure 1. Accuracy and mean anticipation time in DC and the control participants, displayed in ascending order as a function of participants' anticipation. In black: D.C.; in grey: control participants. **(A)** Accuracy: percentage of trials in which participants fixated the target ROI before the hand movement offset. **(B)** Anticipation: participant's mean eye arrival time on the target ROI relative to the hand-movement offset.

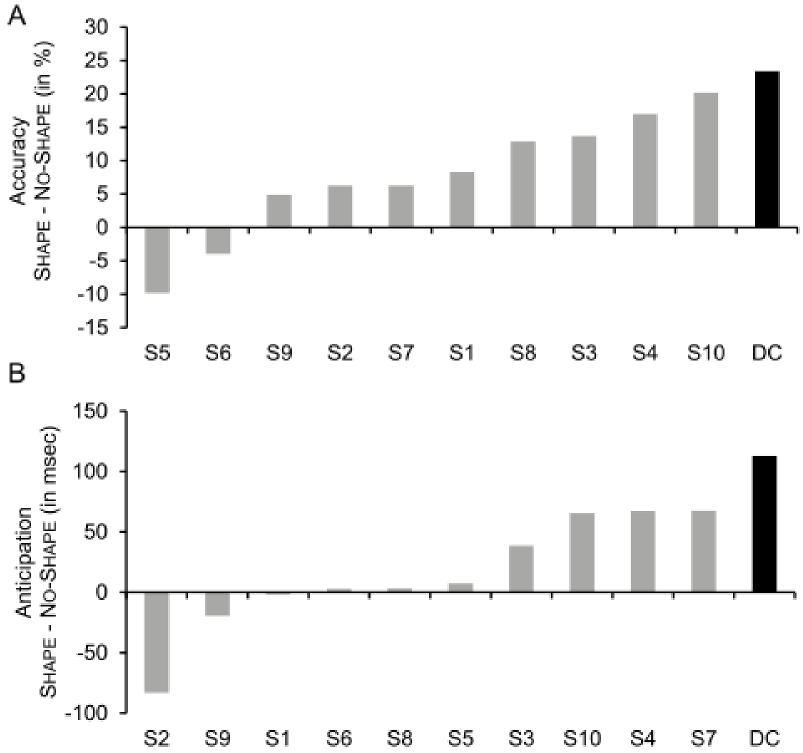

Figure 2. Difference in accuracy and anticipation time for the NO-SHAPE and SHAPE movement conditions in D.C. and each control participant in ascending order as a function of the size of the influence of the shape information. In black: D.C.; in grey: control participants. **(A)** Difference in percentage of trials in which participants fixated the target ROI before the hand movement offset for the SHAPE and NO-SHAPE condition. **(B)** Difference of each participant's mean eye arrival time on the target ROI relative to the hand-movement offset for the SHAPE and NO-SHAPE condition.

Discussion and conclusion

We found that when observing an actor reaching to grasp an object, a person born without upper limbs (D.C.), who therefore has no effector-specific motor representation of the observed action, gazed at the target object before the hand movement offset as often and quickly as typically developed control participants. We also found that, like the control participants, D.C. showed higher accuracy and earlier saccadic movements when he observed a grasping hand than when he observed a reaching hand devoid of a grasping shape. This modulation of predictive eye movements by the actor's hand shape was previously considered to be an index of the crucial role of motor simulation in predictive eye movements during action observation (Ambrosini et al., 2011; Costantini et al., 2013; Falck-Ytter, 2012).

These findings cannot easily be dismissed by appeals to a lack of power or sensitivity. We used a design that has previously been shown to be very sensitive in detecting even subtle impairments of predictive eye movements (Ambrosini et al., 2012; Costantini et al., 2013), and, if anything, D.C. had a tendency to anticipate more quickly, more often, and to be more influenced by the presence of the actor's hand shape than typically developed control participants. Furthermore, the Bayes factors, which unlike the traditional p values allow quantifying the evidence for the null hypothesis, provided strong support in favor of the hypothesis that D.C.'s eye movements were not less predictive or less influenced by the actor's hand shape than the controls' eye movements.

Our findings thus constitute existence proof that it is *possible* to account for typical predictive eye movements, and to explain the modulation of eye movements by actor's hand shape information) *without* appealing to the concept of effector-specific motor simulation. As a corollary, our findings suggest that typical predictive eye movements during action observation can rely exclusively on visual and inferential processes unaided by motor simulation (Eshuis et al., 2009). According to this hypothesis, when an action is perceived, a visuo-perceptual analysis of the observed scene (i.e., of the actor's body shape and motion, objects, context) provides a visual description of the action, which serves as input to an inferential system that makes use of a variety of information and knowledge ("internal models") about the actor, the environment, and the observed action accumulated over previous experience to predict the most likely outcome of the observed action (e.g., Eshuis et al., 2009).

An alternative interpretation of our findings is that predictive eye movements during action observation rely on a kind of motor simulation that is not effector specific. On this account, D.C.'s eye movements would be driven by his covertly imitating the observed hand movements with his lower limbs. However, the very different skeletal and muscular features and degrees of freedom of the hands and feet and fingers and toes make it virtually impossible to imitate the hand movements used in our experiment with the feet. The feet allow neither grasping a large object with a palm power grip nor grasping of a small object with precision grip opposing two fingertips. It is unclear, therefore, how a mechanism of direct, observation-execution matching could underlie D.C.'s predictive eye movements if the observed motor act, for instance grasping a large object with a palm power grip, has never been executed.

In conclusion, our findings challenge the currently received explanation of predictive eye movements during action observation which holds that they cannot be achieved by visual and inferential processes alone but require a direct mapping process of the observed action onto the motor representation of the same body movements in the observer's brain. Of course, it remains an open question the extent to which these results generalize to typically developed participants and to other types of stimuli (e.g., degraded stimuli or stimuli in first person perspective). It is possible, for instance, that predictive eye movements could rely on visual and inferential computations in D.C. but require motor simulation in the typically developed participants. However, as discussed in the introduction, there seem to be currently no compelling evidence supporting this view. In this context, our finding underscores at the very least the need for a shift in the burden of proof relative to the question of the role of motor processes in predictive eye movements during action observation.

Acknowledgements

We are grateful to Marcello Costantini who provided the experimental material and to Angelika Lingnau for her help with the eye tracker. This research was supported by the Fondazione Cassa di Risparmio di Trento e Rovereto and the Provincia Autonoma di Trento.

References

Ambrosini, E., Costantini, M., & Sinigaglia, C. (2011). Grasping with the eyes. *Journal of Neurophysiology, 106*, 1437–1442.

Ambrosini, E., Reddy, V., de Looper, A., Costantini, M., Lopez, B., & Sinigaglia, C. (2013). Looking ahead: anticipatory gaze and motor ability in infancy. *PloS One, 8*, e67916.

Ambrosini, E., Sinigaglia, C., & Costantini, M. (2012). Tie my hands, tie my eyes. *Journal of Experimental Psychology, 38*, 263–266.

Astafiev, S. V., Stanley, C. M., Shulman, G. L., & Corbetta, M. (2004). Extrastriate body area in human occipital cortex responds to the performance of motor actions. Nature neuroscience, 7, 542-548.

Buccino, G., Vogt, S., Ritzi, 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.

Cannon, E. N. & Woodward, A. L. (2008). Action anticipation and interference: a test of prospective gaze. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.). Proc. 30th Annual Conf. Cogn. Sci. Soc. (pp. 981–984). Austin, TX: Cognitive Science Society.

Costantini, M., Ambrosini, E., Cardellicchio, P., & Sinigaglia, C. (2013). How your hand drives my eyes. *Social Cognitive and Affective Neuroscience, 9*, 705-11.

Crawford, J. R., & Garthwaite, P. H. (2007). Comparison of a single case to a control or normative sample in neuropsychology: Development of a Bayesian approach. *Cognitive Neuropsychology, 24*, 343-372.

Crawford, J. R. & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *Clinical Neuropsychologist, 12*, 482-486.

Elsner, C., D'Ausilio, A., Gredeback, G., Falck-Ytter, T., & Fadiga, L. (2013). The motor cortex is causally related to predictive eye movements during action observation. *Neuropsychologia, 51*, 488–492.

Eshuis, R., Coventry, K. R., & Vulchanova, M. (2009). Predictive eye movements are driven by goals, not by the mirror neuron system. *Psychological Science, 20*, 438–40.

Falck-Ytter, T. (2012). Predicting other people's action goals with low-level motor information. *Journal of Neurophysiology, 107*, 2923–5.

Falck-Ytter, T., Gredebäck, G., & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience, 9*, 878–879.

Flanagan, J. R. & Johansson, R. S. (2003). Action plans used in action observation. *Nature, 424*, 769-771.

Hornof, A. J., & Halverson, T. (2002). Cleaning up systematic error in eye-tracking data by using required fixation locations. *Behavior Research Methods, Instruments, & Computers, 34*, 592-604.

Jeffreys, H. (1961). Theory of probability (3 rd Ed.). Oxford, UK: Oxford University Press.

Moreau, D. (2013). Motor expertise modulates movement processing in working memory. *Acta psychologica, 142*, 356-361.

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs. R package version 0.9.10-1. Retrieved from http://CRAN.R-project.org/package=BayesFactor

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*, 356-374.

Southgate, V. (2013). Do infants provide evidence that the mirror system is involved in action understanding? *Consciousness and Cognition, 22*, 1114-1121.

Vannuscorps, G. & Caramazza, A. (2016). Impaired short-term memory for hand postures in individuals born without hands. *Cortex, 83*, 136-138.

Vannuscorps, G., Andres, M., & Pillon, A. (2013). When does action comprehension need motor involvement? Evidence from upper limb dysplasia. *Cognitive Neuropsychology, 30*, 253-283.

Vannuscorps, G., Andres, M., & Pillon, A. (2014). Is motor knowledge part and parcel of the concept of manipulable artifacts ? Clues from a case of upper limb aplasia. *Brain and Cognition, 84*, 132-140.

Vannuscorps, G., Pillon, A., & Andres, M. (2012) The effect of biomechanical constraints in the hand laterality judgment task: where does it come from? *Frontiers in Human Neuroscience, 6*:299.