

# The eye in hand: predicting others' behavior by integrating multiple sources of information

Ettore Ambrosini,<sup>1,2</sup> Giovanni Pezzulo,<sup>3</sup> and Marcello Costantini<sup>1,2,4</sup>

<sup>1</sup>Laboratory of Neuropsychology and Cognitive Neuroscience, Department of Neuroscience, Imaging and Cognitive Sciences, University "G. d'Annunzio," Chieti, Italy; <sup>2</sup>Institute for Advanced Biomedical Technologies, University "G. d'Annunzio," Chieti, Italy; <sup>3</sup>Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy; and <sup>4</sup>Mind, Brain Imaging and Neuroethics, Institute of Mental Health Research, University of Ottawa, Ottawa, Ontario, Canada

Submitted 25 June 2014; accepted in final form 7 January 2015

**Ambrosini E, Pezzulo G, Costantini M.** The eye in hand: predicting others' behavior by integrating multiple sources of information. *J Neurophysiol* 113: 2271–2279, 2015. First published January 7, 2015; doi:10.1152/jn.00464.2014.—The ability to predict the outcome of other beings' actions confers significant adaptive advantages. Experiments have assessed that human action observation can use multiple information sources, but it is currently unknown how they are integrated and how conflicts between them are resolved. To address this issue, we designed an action observation paradigm requiring the integration of multiple, potentially conflicting sources of evidence about the action target: the actor's gaze direction, hand preshape, and arm trajectory, and their availability and relative uncertainty in time. In two experiments, we analyzed participants' action prediction ability by using eye tracking and behavioral measures. The results show that the information provided by the actor's gaze affected participants' explicit predictions. However, results also show that gaze information was disregarded as soon as information on the actor's hand preshape was available, and this latter information source had widespread effects on participants' prediction ability. Furthermore, as the action unfolded in time, participants relied increasingly more on the arm movement source, showing sensitivity to its increasing informativeness. Therefore, the results suggest that the brain forms a robust estimate of the actor's motor intention by integrating multiple sources of information. However, when informative motor cues such as a preshaped hand with a given grip are available and might help in selecting action targets, people tend to capitalize on such motor cues, thus turning out to be more accurate and fast in inferring the object to be manipulated by the other's hand.

action prediction; information integration; gaze direction; hand preshape; arm trajectory

IMAGINE BEING A GOALKEEPER facing a penalty kick. The kicker is approaching the ball from the left while gazing at the corner on your right. Where will you dive? You eventually decide to dive on your right, following the kicker's gaze, but this was a deceptive cue: the kicker kicks to your left, scoring a goal. The ability to predict the outcome of other beings' actions allows humans to adjust their own behavioral output, providing them with a powerful social advantage (e.g., Frith 2007) but also letting them send deceptive cues to score goals (Tomeo et al. 2013). This often occurs effortlessly in everyday life but requires complex computations to solve an ill-posed, inductive problem: the agent's goal and/or intentions are reconstructed from the incoming flow of sensory information providing

multiple (often ambiguous or contradictory) sources of evidence about the agent's goal, such as his gaze direction, arm trajectory, and hand preshape during a reach-to-grasp action (e.g., Manera et al. 2011; Rotman et al. 2006; Sartori et al. 2011). The contribution of each source of information in reconstructing the agent's goal or intentions is still unknown.

To date there is evidence showing that hand preshape is a powerful cue in the understanding of others' actions. In a series of studies (Ambrosini et al. 2011, 2012, 2013; Costantini et al. 2012a, 2012b, 2013) we recorded eye movements while participants observed an actor reaching for and grasping one of two objects requiring two different kinds of grip to be picked up (i.e., precision grip or whole hand prehension). In a control condition, the actor merely reached for and touched one of the two objects without preshaping his hand according to the target features. Results showed that proactive eye movements were faster and more accurate in grabbing the target object when participants observed an actually grasping hand than when they observed a mere touching hand devoid of any target-related preshaping.

Nonetheless, gaze is also considered a powerful cue in the understanding of others' actions. In everyday life we use eye movements to grab and direct the attention of others. Also, we infer the intention of others to act upon objects on the basis of observed eye movements alone (e.g., Castiello 2003; Pierno et al. 2008). This suggests that gaze may be an important cue from which motor intentions of others can be inferred.

In two action observation experiments, we investigated 1) which source of information participants value the most and 2) whether there are differences in how these values are updated during the unfolding of the action, reflecting a sensitivity for the informativeness of the sources (see below). In a first action observation experiment, participants observed an actor's arm movement toward one of two objects requiring different kinds of grip to be picked up (i.e., precision or whole hand grip). In the video stimuli, different sources of evidence about the actor's goal, such as her gaze direction, arm trajectory, and hand preshape, were made available as in natural context. In particular, the actor's gaze direction was made available first, even before the beginning of the actor's action, and only successively did arm trajectory and hand preshape become available. Moreover, we created a conflict between the gaze and preshape information sources by orthogonally manipulating their congruence with the actual target object. We evaluated participants' implicit action prediction ability, as assessed by their predictive gaze behavior and pupillary responses

Address for reprint requests and other correspondence: E. Ambrosini, Dept. of Neurosciences: SNPSRR, Università degli Studi di Padova, Via Giustiniani 5, 35128 Padua, Italy (e-mail: etttore.ambrosini@gmail.com).

during the observation of goal-directed arm movements. In a second experiment, we asked participants to explicitly try to guess the target of the actor's action by performing a mouse movement toward the selected label denoting the action target. In this experiment we also manipulated the amount of total information provided to observers by showing participants different portions of the actor's action, ranging from only 100 to 600 ms.

In agreement with our previous study, we expect that observers can immediately rely on the motor cue provided by the actor's hand preshape to predict the goal of the observed action. But what if additional sources of evidence are available other than the hand preshape, i.e., gaze? Will these information sources be taken into account during the action processing to anticipate the action goal?

Our study has two peculiarities compared with most information integration studies. First, because human action observation is a dynamical task, different sources of evidence are available at different time intervals. The gaze direction is available first, and only successively do the motor cues (i.e., the arm trajectory and hand preshape) become progressively available, so their informativeness can be considered to increase during the course of the action. Second, both the gaze and preshape information are not reliable cues of the final movement, because they correctly cue the agent's goal on 50% of cases. This procedure permits studying if subjects rely more on information provided by motor cues when longer portions of videos are shown, showing a sensitivity to the amount of information the source carries rather than a mere preference for a source over the others.

## MATERIALS AND METHODS

### Experiment 1

**Participants.** Sixteen participants took part in *experiment 1* (10 females, mean age = 23.25 yr, SD = 3.55 yr). All participants were right-handed according to self-report, were naive as to the purpose of the experiment, and had normal or corrected-to-normal visual acuity. Participants provided informed consent before data collection. The procedures were approved by the Ethical Committee of "G. d'Annunzio" University, Chieti, Italy, and were in accordance with the ethical standards of the Declaration of Helsinki.

**Apparatus and stimuli.** Participants were comfortably seated in a chair in front of a 17-in. LCD computer monitor (resolution: 1,024 × 768 pixels; refresh rate: 60 Hz). Their chin and forehead were stabilized by means of a headrest to reduce movement artifacts and to maintain a distance of 57 cm between the participant's eye and the computer monitor. An infrared video-based eye-tracking device (RK-826PCI pupil/corneal tracking system; ISCAN, Burlington, MA), mounted below the monitor, recorded the pupil size and gaze position of the right eye at 120 Hz. The experiment was controlled by a Pentium personal computer using customized software (developed by Gaspare Galati, Department of Psychology, Sapienza University, Rome, Italy; see Galati et al. 2008) implemented in MATLAB (The MathWorks, Natick, MA) using Cogent 2000 (developed at the Leopold Muller Functional Imaging Laboratory and the Institute of Cognitive Neuroscience, University College London, London, UK) and Cogent Graphics (developed by John Romaya, Laboratory of Neurobiology, Wellcome Department of Imaging Neuroscience, University College London).

The experimental stimuli consisted of videos (AVI format; 30 frames/s; 640 × 480 pixels) presented at the center of the computer monitor. They showed from the front view a female actor performing

an unpredictable reaching movement toward either a small or a large tomato (targets), both located on a table at a distance of ~50 cm from her torso and ~20 cm apart from each other. The small and large targets subtended  $1.34^\circ \times 1.27^\circ$  and  $3.60^\circ \times 2.96^\circ$ , respectively, and were distributed symmetrically about the vertical midline (see Fig. 1), according to two different object layouts obtained by switching the object locations. There were a total of 16 different videos (2 object layouts × 2 targets × 2 gaze directions × 2 hand preshapes), and thus four different videos were shown (8 times each) for each of the four experimental conditions (see below), resulting in a total of 128 trials.

All the videos started with the actor looking at one of the objects for 1,000 ms (the large and the small object were fixated equally often). During this time, her right hand was resting on the table immediately in front of her torso and a black fixation cross was presented in the center of the screen (Fig. 1, fixation phase). Next, while maintaining the fixation, the actor started moving her hand toward one of the targets, which was independent of the fixated object (Fig. 1, movement phase). In other words, in half of the trials the fixated and the reached object were the same object (Fig. 1, *A* and *B*), whereas in the other half they were different objects (Fig. 1, *C* and *D*). Moreover, during the reaching movement, in half of the trials the actor's hand shaped a precision grip, whereas in the other half her hand shaped a whole hand grip,<sup>1</sup> making them visible as soon as the reaching movement started. The actor was instructed not to move other body parts during the fixation and movement phases to hide movement cues that indicate movement preparation and to perform her reach-to-grasp movements as naturally and as smoothly as possible. The videos showed the entire actor's arm movement, i.e., from the earliest detectable movement of the hand to the full hand-object contact (movement phase), lasting ~1,250 ms ( $1,262 \pm 114$  ms, mean ± SD). The final 500 ms of the video then consisted of the last frame of the movement phase, which was shown as still (contact phase) (see Fig. 1). In total, each stimulus lasted ~2,750 ms.

To sum up, for each actor's hand movement toward a given target, participants were presented with videos belonging to four conditions based on the congruence between the information conveyed by both the actor's gaze and hand preshape and the actual target of the actor's hand movement, which was determined by the hand trajectory (i.e., the hand trajectory was always a congruent cue of the reach target). Therefore, the four experimental conditions were 1) gaze congruent-preshape congruent, in which both gaze and preshape cued the same object as the target of the actor's action; 2) gaze congruent-preshape incongruent, in which the actor's gaze was informative of the actor's actual target whereas the preshape was not; 3) gaze incongruent-preshape congruent, in which the actor's hand preshape was informative of the actor's actual target whereas the gaze was not; and 4) gaze incongruent-preshape incongruent, in which both gaze and preshape were misleading in cueing the actual target (see Fig. 1).

**Procedure.** To minimize participants' fatigue, the experiment was divided into two blocks during which we recorded participants' gaze position and pupil size. At the beginning of each block, participants' gaze position was calibrated using a standard nine-point calibration procedure. Each experimental block consisted of 64 trials (16 repetitions for each of the 4 experimental conditions: gaze congruent-preshape congruent, gaze congruent-preshape incongruent, gaze incongruent-preshape congruent, and gaze incongruent-preshape incongruent) and lasted less than 6 min. The order of trials within each block was randomized, and the order of blocks was counterbalanced across participants. Each trial began with the presentation of the

<sup>1</sup> As in our previous works, we chose to use the precision and whole hand grasp types because they are easily distinguishable visually, having distinctive kinematic (e.g., Gentilucci et al. 1991; Jeannerod 1988) and neural (e.g., Begliomini et al. 2007a; Ehrsson et al. 2000) signatures. However, unlike in our previous works, we did not investigate potential influences of the type of grasp/target object because this would be outside the scope of the present study.

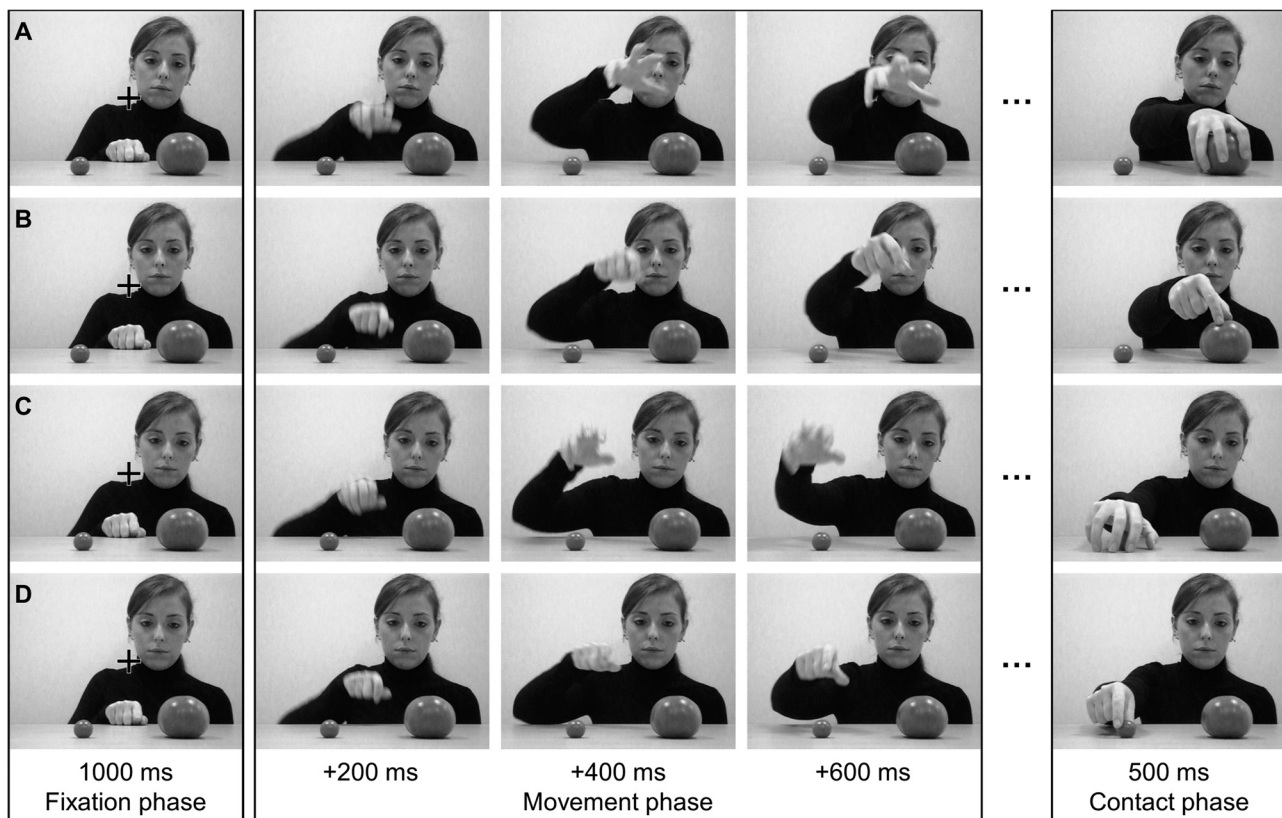


Fig. 1. Exemplar of movement kinematics in each experimental condition: gaze congruent-preshape congruent (A); gaze congruent-preshape incongruent (B); gaze incongruent-preshape incongruent (C); and gaze incongruent-preshape congruent (D). The images shown at far right were used in *experiment 1* only. Note that only 4 of the 16 different videos used are shown, i.e., those for 1 of the 2 object layouts and for 1 of the 2 gaze directions (see MATERIALS AND METHODS for details).

stimulus video, and the participants were asked to move their gaze on the fixation cross at the center of the screen until its disappearance and then to simply watch the video. During the intertrial interval (2,500 ms), a white fixation cross on a gray background indicated the blinking period during which the participants were allowed (and recommended) to blink. Indeed, we asked participants to try to restrict eye blinks to the blinking phase at the end of the trial to reduce blinking and artifacts during stimulus presentation and thus minimize the number of excluded trials.

**Data analysis.** We analyzed participants' gaze position recorded during the observation of the video stimuli using an I-VT (velocity-threshold identification) algorithm written with MATLAB (The Math-Works, Natick, MA) that automatically detected saccades by means of both a velocity and a temporal threshold (point-to-point velocity of the gaze trace  $> 50^\circ/\text{s}$  for 2 consecutive samples). This algorithm was modified from Salvucci and Goldberg (2000) by adding a temporal criterion to mitigate the instrument noise and prevent saccade mis-identifications (Ambrosini et al. 2011). For each trial, we created two areas of interest (AOI), covering the fixation cross (fixation AOI) and the intended target (target AOI). The target AOI was actually  $0.2^\circ$  larger than the real stimulus to compensate for noise in the eye-tracking system.

A total of 1,920 trials were recorded ( $64 \text{ trials} \times 2 \text{ blocks} \times 15 \text{ participants}$ ). All the analyses of participants' gaze behavior were performed by considering only trials in which participants exhibited a target-directed gaze behavior; that is, trials in which participants did not fixate the fixation AOI at the beginning of the movement phase (4.69% of the recorded trials) or in which they did not make a saccade to the target AOI at any point before the end of the video (26.67% of the remaining trials) were excluded and not further analyzed. Therefore, we did not consider as predictive the occasional gaze shifts to the

objects before the actor had started to move. Note that the percentage of trials in which participants did not fixate the target was not dissimilar to that found in our previous studies (Ambrosini et al. 2011, 2012, 2013; Costantini et al. 2012a, 2012b) using similar tasks (range = 18–31%; mean = 24%). Moreover, it should be stressed that our aim was to investigate participants' action prediction ability, so for this reason we chose to selectively analyze trials in which participants exhibited a target-directed gaze behavior. For each remaining trial, we calculated the arrival time of the gaze on the target AOI (gaze arrival time) as a dependent variable to assess the predictive nature of participants' gaze behavior, i.e., their ability to anticipate with the eyes the goal of the observed action. The gaze arrival time was computed by subtracting the time when participants first looked inside the target AOI from the hand-object contact time (i.e., the end of the movement phase). Therefore, if the participant's gaze arrived at the target AOI before the end of the actor's action, the trial was regarded as predictive and the gaze arrival time took a negative score. Our choice about the threshold for gaze anticipations was quite conservative. Indeed, in line with prior studies on action understanding and goal anticipation (e.g., Falck-Ytter et al. 2006; see also Flanagan and Johansson 2003), we chose a temporal threshold of 0 ms instead of a more liberal criterion incorporating a 200-ms reaction time in anticipations (e.g., Gredeback et al. 2009, 2010). Therefore, our estimates of participants' goal anticipations would heavily underestimate the actual degree of their gaze proactivity.

Regarding the pupil size data, the analysis was conducted in trials in which participants' gaze was within the fixation AOI during the last 200 ms of the fixation phase (95.4% of the recorded trials). We developed an in-house algorithm, written with MATLAB (The Math-Works), to remove blinks as well as other minor artifacts (Montefinese et al. 2013). Blinks were identified as sudden large changes in vertical



pupil diameter and were filled in by cubic spline interpolation. The percentage of interpolated samples (mean = 4.09%) was not systematically distributed across experimental conditions [ $F(1,14) \leq 0.22$ ,  $P \geq 0.65$ ], and no single trial presented a high number of interpolated points (>30%). Resulting pupillary data were then smoothed using an unweighted seven-point moving median filter to remove instrumental noise. Constant fluctuation in pupil size over time and interindividual variations were controlled by computing an index that quantifies the percentage of change in pupil diameter (PDC) due to the processing of the video stimuli compared with a baseline (prestimulus) pupil diameter for each trial. This measure was computed for each sample during the movement and contact phases by subtracting the pupil diameter from the baseline pupil diameter (i.e., the mean pupil size during the last 200 ms of the fixation phase), dividing by the baseline pupil diameter, and multiplying by 100. In this manner, pupil size changes were independent from initial pupil size and comparable between participants.

The effect of the experimental manipulation on the dependent variables described above was assessed by conducting linear mixed effects modeling as implemented by the function *lmer* from the *lme4* library (Bates et al. 2012) in R (version 2.15.2; R Core Team, 2012). This approach has several advantages over traditional general linear model analyses (such as repeated-measures ANOVA) that made it suitable for the present data. First, unlike general linear models, mixed-effects models are very robust with respect to missing data and unbalanced data sets (Baayen et al. 2008; Quené and van den Bergh 2008). Moreover, because mixed-effects model analyses are conducted on trial-level data (i.e., they do not require prior averaging across participants, whereas by-items multiple regression models do), they offer the possibility of preserving and taking into account any variability across individuals, thus increasing the accuracy and generalizability of the parameter estimate. This allowed us to account for random and fixed effects at the within- and between-subject levels, providing more efficient estimates of the experimental effects and a better protection against capitalization on chance, or type I error (Baayen et al. 2008; Quené and van den Bergh 2008).

The experimental effects were incongruence (or violation) effects and a linear function of the time throughout the experiment. The incongruence effects, modeled as dummy variables (0 = congruent, 1 = incongruent), corresponded to a  $2 \times 2$  factorial design with two main effects corresponding to our experimental manipulations, namely, the incongruence of gaze and hand preshape with the actual target. The effect of the time throughout the experiment (i.e., the factor time) was modeled by a parameter representing the trial number vector (zero-centered to remove the possible spurious correlation between the by-subjects random intercepts and slopes); this main effect accounts for potential confounding longitudinal effects of fatigue or familiarization across participants. This design allowed us to look for the effects of main theoretical interest, that is, the interactions and main effects among the experimental conditions and how these effects depended on the participants' experience as the reliability of the various cues became apparent (i.e., interaction of congruencies with time).

We determined the simplest best (final) linear mixed-effect models to fit our dependent variables by using a log-likelihood ratio test (for a detailed description of the procedure, see Montefinese et al. 2014) according to standard procedures (e.g., Baayen et al. 2008; Quené and van den Bergh 2008). Specifically, we started the model-building process by modeling the random part of the model, which include in all the cases three parameters for the residual error, the random effect of subjects, and the by-subjects random slopes for time. We then tested for the inclusion of the parameter for the linear function of time to partial out the effect of this potentially confounding variable. Finally, we tested for the inclusion of parameters for the fixed effects of interest, namely, the full-factorial combination of the gaze and preshape factors and the linear function of time. These fixed effects account for our predictions. Unless otherwise specified, the fixed part

of the resulting final model included five parameters for the fixed effects of intercept, time, gaze (congruent vs. incongruent), preshape (congruent vs. incongruent), and the gaze by preshape interaction. After this model-building procedure, the statistical significance of the fixed effects included in the final model was assessed as detailed below.

For each continuous dependent variable, we fit the final model after excluding outliers, which were identified as observations for which the standardized residual exceeded the value of  $\pm 3$  (always <2.5% of total observations). For fixed effects, we reported the estimated coefficient (*b*), standard error (SE), and *t* values for each parameter included in the final model. In addition, we reported the *P* values ( $P_{\text{MCMC}}$ ) and upper and lower highest posteriori density intervals (HPD<sub>95%</sub>) estimated on the basis of the posterior distribution of the corresponding parameters, obtained through Markov Chain Monte Carlo (MCMC) sampling (10,000 samples) supported by the *pvals.fnc* function of the language R package (version 1.4; Baayen et al. 2008).

### Experiment 2

**Participants.** Fifteen participants took part in *experiment 2* (10 females, mean age = 22.33 yr, SD = 2.53 yr). All participants were right-handed according to self-report, were naive as to the purpose of the experiment, and had normal or corrected-to-normal visual acuity. Participants provided informed consent before data collection. The procedures were approved by the Ethical Committee of "G. d'Annunzio" University, Chieti, and were in accordance with the ethical standards of the Declaration of Helsinki.

**Apparatus and stimuli.** The experimental stimuli were the same videos used in *experiment 1*, but in this case we constructed six different versions of each video by varying the duration of the movement phase so that the videos ended either 3, 6, 9, 12, 15, or 18 frames (at 30 Hz; 100–600 ms) after the actor's hand started moving from its resting position toward one of the two objects. It should be noted that since the videos had slightly variable total durations, slightly different portions of the reaching movement were shown for each video. However, both the fact that four different videos were presented for each condition and the lack of significant correlations between the total duration of the videos and the dependent variables [respectively,  $r(16) = -0.21$ , 0.27, and 0.03 for accuracy, response times, and area under the curve (see *Data analysis* below), all  $P > 0.31$ ] would suggest that this potential drawback did not affect the validity of our results. There were a total of 96 different videos (2 object layouts  $\times$  2 targets  $\times$  2 gaze directions  $\times$  2 hand preshapes  $\times$  6 durations). The videos were presented at the center of a 15.6-in. monitor (resolution: 1,366  $\times$  768 pixels, refresh rate: 60 Hz) placed 57 cm in front of the participant's eye.

The presentation of the stimuli and the recording of the participants' responses were controlled using the MouseTracker software package, which is a freely available, self-contained application developed specifically for the design, recording, and analysis of mouse-tracking experiments (Freeman and Ambady 2010).

**Procedure.** Before each trial began, a small box labeled "START" was shown at the bottom center of the screen. After the participants clicked the start box to initiate the trial, two response boxes labeled "POMODORO" and "PACHINO" (i.e., the Italian words representing the large and the small tomato, respectively) appeared at the top left and top right corners of the screen and a random video was presented at the center of the screen. The participants were required to carefully watch the video and try to guess which object was the target of the actor's hand movement, and to do this as quickly and accurately as possible. To provide their response, participants moved the mouse cursor forward from the starting position toward the top of the screen to click on the chosen response box. Meanwhile, the streaming *x*- and *y*-coordinates of the mouse were recorded at a sampling rate of  $\sim 70$  Hz. To ensure mouse trajectories were online with decision processes, we asked participants to begin initiating movement as early as possi-

Table 1. Estimated parameters and statistics of linear mixed-effects modeling of gaze arrival times

Fixed Effects	<i>b</i>	SE	<i>t</i>	HPD95 <sub>lower</sub>	HPD95 <sub>upper</sub>	<i>P</i> <sub>MCMC</sub>
Intercept	-483.9449	53.8189	-8.992	-571.0350	-389.7945	0.0001
Time	-0.0792	0.4407	-0.180	-1.0170	0.9538	0.8664
Preshape	62.0176	2.2241	3.067	23.1020	101.5306	0.0016
Gaze	-37.6387	19.9686	-1.885	-78.0750	0.0245	0.0578
Preshape × gaze	66.7474	28.7615	2.321	8.3090	12.5215	0.0208

Values are the coefficient (*b*), standard error (SE), and *t* values for each parameter. HPD95<sub>lower</sub> and HPD95<sub>upper</sub>, lower and upper intervals of highest posterior density at 95%, respectively; *P*<sub>MCMC</sub>, *P* values estimated using Markov Chain Monte Carlo sampling.

ble (note that during the fixation phase, the mouse was not allowed to move from the start location). Once the participant clicked on the response box, the start box appeared for the participant to initiate the next trial.

Each of the 96 videos was presented 5 times in the experiment, and thus participants performed 480 trials. After half of the trials had been completed, the horizontal location of the two response labels was flipped (which response appeared on the left/right top corner during the first block was counterbalanced across participants). Trials were presented in randomized order.

**Data analysis.** Trials in which participants did not respond within a 5,000-ms time window were discarded (60 of 7,680 recorded trials, corresponding to 0.78%). Dependent variables calculated on remaining trials included accuracy, response times (RTs), and mouse trajectory data, all recorded by MouseTracker. Accuracy was a binary index measuring whether participants provided a correct response in a given trial. RTs quantified the time elapsed (in ms) between the click on the start button (triggering the presentation of the stimulus video) and the click on the response button. RTs were log-transformed to mitigate the influence of nonnormal distribution and skewed data.

Regarding the mouse-tracking data, we first transformed mouse trajectories according to standard procedures (Freeman and Ambady 2010). In particular, all trajectories were rescaled into a standard coordinate space (top left = [-1, 1.5]; bottom right = [1, 0]) and flipped along the *x*-axis such that they were directed to the top right corner. Moreover, all trajectories were time-normalized into 101 time steps with the use of linear interpolation to permit averaging of their full length across multiple trials. To obtain a trial-by-trial index of the trajectory's attraction toward the nonselected response label (indexing how much that response was simultaneously active), we computed a summary measure called area under the curve (AUC), which is a common index for assessing response competition (i.e., larger positive AUC values indicate greater response competition and more difficulty in making a decision). This index is calculated as the area between the actual trajectory and its idealized trajectory (a straight line between each trajectory's start and endpoints) out of all time steps, quantifying how far a trajectory deviates toward the unselected option before the participant ultimately selects the chosen option. We also computed another commonly used measure of response competition, maximum deviation (MD), but did not include it here because the analysis of this measure yielded results similar to those of AUC. We indeed chose to report AUC because it is a more global and stable measure of the trajectory deviation compared with MD, which is calculated on the basis of a single point of the mouse movement trajectory.

The effect of the experimental manipulation on the dependent variables described above was assessed by conducting linear mixed-effects modeling as described for *experiment 1*. In *experiment 2*, there was an additional predictor encoding the duration or amount of visual information available to each participant. Unless otherwise specified, the fixed part of the final model included nine parameters for the fixed effects of intercept, time, gaze (congruent vs. incongruent), preshape (congruent vs. incongruent), duration of the video movement phase (6 levels, from 100 to 600 ms), and the two-way and three-way interactions involving the latter three factors. Moreover, since accuracy is a binary dependent variable, we fitted it with a generalized linear mixed

model using the *lmer* function again, but in this case selecting the binomial distribution and the logistic link function. Note that in this case we provided *b*, SE, *z*, and *P* values for each parameter.

## RESULTS

### Experiment 1

**Gaze arrival times.** The analysis conducted on gaze arrival times (see Table 1) revealed the significant main effect of preshape, showing that participants were earlier in gazing at the intended target of the actor's hand movement when the actor's hand preshape congruently cued her goal (*b* = 62.018, SE = 20.224, *t* = 3.07, HPD<sub>95%</sub> = 23.102 to 101.531, *P*<sub>MCMC</sub> = 0.002). The interaction gaze by preshape was also significant (*b* = 66.747, SE = 28.762, *t* = 2.32, HPD<sub>95%</sub> = 8.309 to 120.522, *P*<sub>MCMC</sub> = 0.021), indicating that participants gazed the target AOI later when both the sources of information were misleading (i.e., in the gaze incongruent-preshape incongruent condition; see Fig. 2) compared with all the other experimental conditions. In addition, when the hand preshape was congruent with the intended target of the actor's hand movement, participants' gaze behavior was faster and accurate regardless of the information provided by the actor's gaze (see Fig. 2). No other main effects or interactions reached the significance level (see Table 1).

**Mean pupil dilation change.** The model-building procedure revealed that the inclusion of neither the parameter for the main effect of gaze nor that for the gaze by preshape interaction was justified [ $\chi^2(1) < 0.64$ , *P* > 0.42, and  $\chi^2(2) < 1.02$ , *P* > 0.60, respectively] (see Table 2 for the parameters of the final model). The analysis revealed a significant main effect of preshape, indicating a stronger pupillary response when the

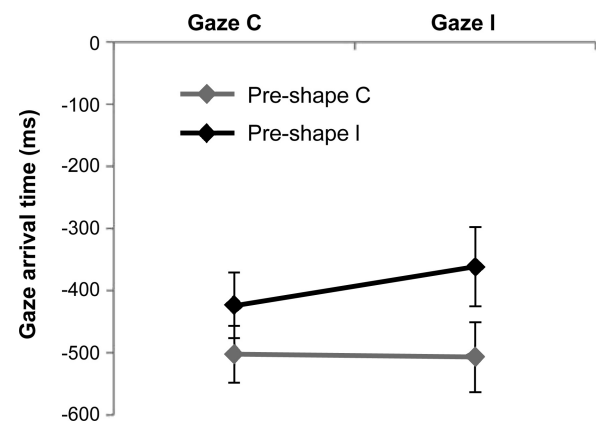


Fig. 2. Participants' mean gaze arrival time in *experiment 1* plotted as a function of the actor's gaze direction and hand preshape. Error bars indicate SE. C, congruent; I, incongruent.

Table 2. Estimated parameters and statistics of linear mixed-effects modeling of mean pupil dilation

Fixed Effects	<i>b</i>	SE	<i>t</i>	HPD95 <sub>lower</sub>	HPD95 <sub>upper</sub>	<i>P</i> <sub>MCMC</sub>
Intercept	1.0904	0.0099	11.35	1.0730	1.1082	0.0001
Time	0.0001	0.0001	0.77	−0.0001	0.0003	0.4732
Preshape	0.0089	0.0030	3.02	0.0031	0.0149	0.0036

actor's hand preshape was a deceptive source of information about the actor's goal ( $b = 0.009$ ,  $SE = 0.003$ ,  $t = 3.02$ ,  $HPD_{95\%} = 0.003$  to  $0.015$ ,  $P_{MCMC} = 0.004$ ). The main effect of time was not significant ( $b = 0.0001$ ,  $SE = 0.0001$ ,  $t = 0.77$ ,  $HPD_{95\%} = -0.0001$  to  $0.0003$ ,  $P_{MCMC} = 0.473$ ).

### Experiment 2

**Accuracy.** Table 3 shows the summary of the final model. Note that in this case the final model also included a parameter for the time by gaze interaction, which significantly improved the model fit [ $\chi^2(1) = 23.78$ ,  $P < 2 \times 10^{-6}$ ]. The mixed model analysis revealed the significant main effect of time ( $b = -1.56 \times 10^{-3}$ ,  $SE = 7.83 \times 10^{-4}$ ,  $z = -1.991$ ,  $P = 0.0465$ ), showing that, on average, there was a learning effect: participants' accuracy in predicting the actor's goal increased as the experiment ensued. There was also a significant effect of duration, indicating that participants' accuracy in guessing the actor's goal increased as more information was available about the observed hand action ( $b = 0.013$ ,  $SE = 1.42 \times 10^{-3}$ ,  $z = 9.223$ ,  $P < 2 \times 10^{-16}$ ).

The main effect of gaze was also significant, with incongruent actor's gaze direction that caused lower accuracy ( $b = -1.705$ ,  $SE = 0.283$ ,  $z = -6.033$ ,  $P < 2 \times 10^{-9}$ ). Moreover, the time by gaze interaction was significant ( $b = 3.44 \times 10^{-3}$ ,  $SE = 5.28 \times 10^{-4}$ ,  $z = 6.512$ ,  $P < 8 \times 10^{-11}$ ), showing that the participants' reliance on information provided by the actor's gaze direction was modulated by learning. In fact, the detrimental effect of the actor's incongruent gaze direction decreased as the task ensued. The duration by preshape interaction was significant ( $b = -7.32 \times 10^{-3}$ ,  $SE = 1.50 \times 10^{-3}$ ,  $z = -4.868$ ,  $P < 2 \times 10^{-6}$ ), showing that the beneficial effect of the preshape congruency, which led to a steeper improvement of participants' accuracy as more information was provided, was abolished for the longest duration, i.e., when the information about the hand trajectory undoubtedly informed participants' about the target of the reach action (accuracy > 95%). This effect was further qualified by a significant duration by preshape by gaze interaction ( $b = 5.06 \times 10^{-3}$ ,  $SE = 1.66 \times 10^{-3}$ ,  $z = 3.045$ ,  $P < 0.003$ , respectively). This higher order

Table 3. Estimated parameters and statistics of generalized linear mixed-effects modeling of accuracy

Fixed Effects	<i>b</i>	SE	<i>z</i>	<i>P</i>
Intercept	0.0639	0.251	0.255	0.7985
Time	−0.0016	0.001	−1.991	0.0465
Duration	0.0131	0.001	9.223	<0.0001
Preshape	0.1938	0.281	0.689	0.4909
Gaze	−1.7047	0.283	−6.033	<0.0001
Gaze time	0.0034	0.001	6.512	<0.0001
Duration × preshape	−0.0073	0.002	−4.868	<0.0001
Duration × gaze	−0.0024	0.002	−1.593	0.1112
Preshape × gaze	−0.4364	0.342	−1.274	0.2025
Duration × preshape × gaze	0.0051	0.002	3.045	0.0023

interaction shows that the information provided by the actor's gaze was modulated by that provided by her hand preshape, because participants' accuracy was higher when both sources correctly cued the targets and was lower in the opposite case. Moreover, participants' accuracy was deeply impacted by gaze information only when hand preshape information was not available, that is, when only 100 or 200 ms of the entire movement was shown. Conversely, the effect of preshape congruency on participants' performance was abolished only for the longest duration, for which the participants' accuracy was at ceiling. No other main effects or interactions were significant (see Table 3).

**Response times.** The analysis performed on RTs revealed the significant main effect of time, showing that, on average, there was a longitudinal familiarization effect (see Table 4). In fact, participants were faster in finalizing the mouse response as the experiment ensued ( $b = -4.79 \times 10^{-4}$ ,  $SE = 1.01 \times 10^{-4}$ ,  $t = -4.75$ ,  $HPD_{95\%} = -0.0007$  to  $-0.0003$ ,  $P_{MCMC} \leq 0.0001$ ). The analyses also revealed a significant main effect of duration, indicating that RTs were faster as more visual detail of the actor's action became available ( $b = -6.05 \times 10^{-4}$ ,  $SE = 3.08 \times 10^{-5}$ ,  $t = -19.64$ ,  $HPD_{95\%} = -0.0007$  to  $-0.0005$ ,  $P_{MCMC} \leq 0.0001$ ).

Moreover, the main effect of gaze was significant ( $b = 0.049$ ,  $SE = 0.020$ ,  $t = 2.49$ ,  $HPD_{95\%} = 0.011$  to  $0.088$ ,  $P_{MCMC} = 0.012$ ), suggesting that participants were slower in responding when the information provided by the actor's gaze misleadingly cued her goal. The duration by preshape interaction was also significant ( $b = 1.18 \times 10^{-4}$ ,  $SE = 4.41 \times 10^{-5}$ ,  $t = 2.68$ ,  $HPD_{95\%} = 0.0001$  to  $0.0002$ ,  $P_{MCMC} = 0.008$ ), showing that the detrimental effect of incongruent preshape on participants' response times increased as the actor's action unfolded. No other main effects or interactions were significant (see Table 4).

**Area under the curve.** The results of the analyses on the mouse-tracking index measuring response competition are shown in Table 5. The main effect of preshape was significant ( $b = 0.031$ ,  $SE = 0.015$ ,  $t = 2.03$ ,  $HPD_{95\%} = 0.001$  to  $0.061$ ,  $P_{MCMC} = 0.044$ ), suggesting that when participants were presented with an incongruent hand preshape, their mouse responses were more attracted by the (unselected) response alternative, that is, by the response erroneously cued by the observed hand preshape. No other main effects or interactions reached the significance level (see Table 5).

### DISCUSSION

In this study we investigated the contribution of gaze and hand preshape in action understanding. In particular, we tried to answer two experimental questions: 1) which source of information (i.e., gaze vs. preshape) participants value the most while observing reach-to-grasp movements and 2) whether these values are fixed, reflecting a static preference for one source over the others, or updated during the unfolding of the



Table 4. *Estimated parameters and statistics of linear mixed-effects modeling of response times*

Fixed Effects	<i>b</i>	SE	<i>t</i>	HPD95 <sub>lower</sub>	HPD95 <sub>upper</sub>	<i>P</i> <sub>MCMC</sub>
Intercept	7.237	0.076	95.01	7.1487	7.3273	0.0001
Time	−4.79E−04	1.01E−04	−4.75	−0.0007	−0.0003	0.0008
Duration	−6.05E−04	3.08E−05	−19.64	−0.0007	−0.0005	0.0001
Preshape	−0.023	0.018	−1.30	−0.0569	0.0144	0.2078
Gaze	0.049	0.020	2.49	0.0112	0.0882	0.0124
Duration × preshape	1.18E−04	4.41E−05	2.68	<0.0001	0.0002	0.0080
Duration × gaze	−3.49E−05	4.72E−05	−0.74	−0.0001	0.0001	0.4532
Preshape × gaze	−0.012	0.029	−0.43	−0.0677	0.0452	0.6696
Duration × preshape × gaze	−5.36E−05	6.84E−05	−0.78	−0.0002	0.0001	0.4438

action, reflecting a sensitivity for the changing availability and informativeness of the sources. In two action observation experiments we assessed participants' prediction of the goal of an actor's arm movement toward one of two objects requiring different kinds of grip to be picked up (i.e., precision or whole hand grip). To test the dynamic interaction among different information sources cueing the actor's goal, namely, gaze direction, hand preshape, and arm trajectory, we made them available with different degrees of reliability at different moments during the videos showing the actor's actions.

Our results show that the actor's gaze direction had an effect on participants' explicit prediction ability. Indeed, when this information misleadingly cued the actor's goal, participants were less accurate and slower in providing the mouse response to express their explicit judgments in *experiment 2*. This result confirms the key role of gaze direction as a crucial information source about others' actions. Indeed, it is a fundamental social cue and plays a pivotal role in social cognition, providing ample information about others' mental and emotional states (Baron-Cohen et al. 2001), allowing one to detect their focus of attention (Nummenmaa and Calder 2009; Ramsey et al. 2011), and automatically triggering an attention shift to the same location (Friesen and Kingstone 1998; but see Ricciardelli et al. 2012; for a review on the influence of gaze processing on object processing, see Becchio et al. 2008). Moreover, in most real-life cases, the actor's gaze direction is sufficient for humans to infer his/her motor intention (Castiello 2003). Interestingly, a study by Pierno et al. (2008) has shown that merely observing someone else's gaze shift toward an object led to the activation of cortical areas known to be involved in processing hand-object interactions. The same study also showed that the activity in the inferior frontal gyrus was modulated by the relationship between the model's gaze and the objects, suggesting that this cortical area has a crucial role in processing not only hand-object but also gaze-object interactions (Pierno et al. 2008).

However, differently from what happens in everyday life, in our paradigm the actor's gaze was unreliable, cueing the correct response in the 50% of the cases. In our case, therefore, a rigid reliance on the actor's gaze would be highly detrimental for participants' performance. The analysis of participants' accuracy in *experiment 2* speaks, indeed, against a rigid reliance on this source of information, because both the hand preshape and the video duration (i.e., amount of available information) modulated the detrimental effect of gaze incongruence on participants' judgments (see Fig. 3). This suggests that gaze was highly influential when no other information about the actor's behavioral intention was provided (e.g., for shorter videos). This is supported by the evidence showing that when the hand preshape correctly cues the actor's goal and/or the video duration increases, the information provided by the gaze decreases (Hudson and Jellema 2011). Accordingly, the results of *experiment 2* show that, for longer videos, the importance lowers for gaze and rises for arm trajectory (which carries increasingly more information relative to the correct goal), especially with 600-ms-long videos, when the impact of arm trajectory information overwhelms gaze and preshape information, abolishing their effects on participants' performance and leading to a ceiling level of accuracy.

Our results also show that the actor's hand preshape had widespread effects on participants' prediction ability, affecting both their predictive eye movements and their mouse responses. Indeed, our results show that participants were much more accurate and fast in gazing at the object to be manipulated by the other's hand when the actor's hand preshape was congruent with the intended target of the actor's hand movement, regardless of the information provided by the actor's gaze. Moreover, the actor's hand preshape was the only information source that affected the kinematic of the participants' mouse responses, attracting them toward the response option

Table 5. *Estimated parameters and statistics of linear mixed-effects modeling of area under the curve*

Fixed Effects	<i>b</i>	SE	<i>t</i>	HPD95 <sub>lower</sub>	HPD95 <sub>upper</sub>	<i>P</i> <sub>MCMC</sub>
Intercept	0.0303	0.0157	1.935	−0.0006	0.0619	0.0586
Time	5.86E−05	3.58E−05	1.634	0	0.0001	0.1260
Duration	6.80E−06	2.63E−05	0.259	0	0.0001	0.8026
Preshape	0.0312	0.0154	2.029	0.0010	0.0612	0.0440
Gaze	0.0012	0.0168	0.073	−0.0313	0.0338	0.9428
Duration × preshape	−2.22E−05	3.78E−05	−0.588	−0.0001	0	0.5656
Duration × gaze	4.71E−06	4.02E−05	0.117	−0.0001	0.0001	0.9080
Preshape × gaze	−0.0472	0.0249	−1.900	−0.0962	0	0.0512
Duration × preshape × gaze	6.26E−05	5.87E−05	1.066	−0.0001	0.0002	0.2796

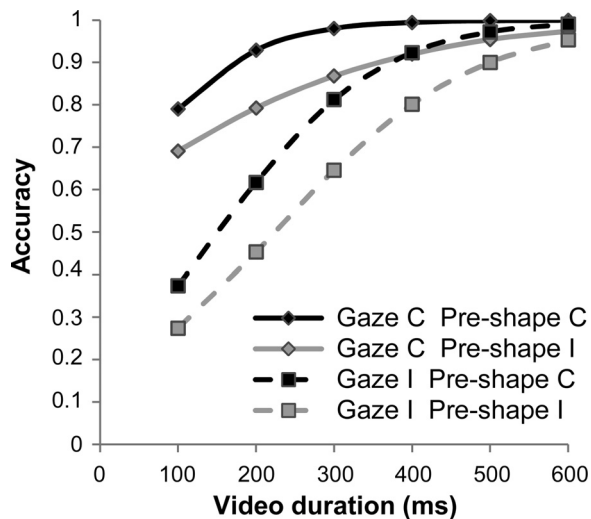


Fig. 3. Participants' accuracy for each video duration in *experiment 2* plotted as a function of the actor's gaze direction and hand preshape.

cued by the actor's hand preshape.<sup>2</sup> These results thus suggest that observing an agent's hand preshape automatically evokes motor representations of the action-object relationship (e.g., Rizzolatti and Sinigaglia 2010; see also Becchio et al. 2012), implying the detection of the potential for successful action outcomes (Bach et al. 2011). Moreover, they extend our previous findings (Ambrosini et al. 2011, 2012, 2013; Costantini et al. 2012a, 2012b, 2013; see also Kanakogi and Itakura 2011) by showing that the agent's hand preshape provides the observer with enough motor cues to anticipate with his/her gaze the target object of the observed action, even when contrasting sources of evidence such as the actor's gaze direction are presented simultaneously. Finally, an interesting result was revealed by the analysis of pupil size, showing that participants' pupillary response during the observation of the actor's action was stronger when her hand preshape misleadingly cued her goal but not when the gaze misleadingly cued her goal. This result suggests that whereas misleading information regarding the hand preshape violated participants' expectancies regarding the flow of observed events (O'Reilly et al. 2013; Preusschoff et al. 2011; see also Gredeback and Melinder 2010), this did not occur when the gaze misleadingly cued the actor's intended target.

Finally, the fact that participants relied more on arm movements as the action unfolded in time (e.g., in longer videos) is in keeping with the idea that multiple sources of evidence can be integrated and weighted depending on their reliability, a principle that has been demonstrated in perceptual (Ernst and Bulthoff 2004) and motor domains (Kording and Wolpert

2006). Our study provides for the first time evidence that similar principles might be at work during action perception, which is compatible with recent proposals that cast it in terms of hierarchical probabilistic inference and predictive coding (Dindo et al. 2011; Friston et al. 2011; Kilner et al. 2007; Pezzulo 2013; Pezzulo et al. 2013). At the same time, our results show systematic biases in the integration process: participants continued using hand preshape as a source of information despite its reliability being fixed at 0.5, as revealed by the fact that it affected both the participants' action prediction ability in *experiment 2* and their predictive gaze behavior in *experiment 1*. Formally speaking, participants behaved in a way that is dictated by their "hyperpriors" (i.e., prior beliefs about precision of a given source that derive from previous experience; Friston 2010), and they fail to update these (hyper)priors during the experiment. Conversely, our results show that the effect of the actor's gaze on the accuracy of participants' explicit judgments decreased as the experiment ensued, suggesting that the participants' reliance on this information source was modulated by learning. Taken together, our results speak to a difference between the way gaze and hand preshape are integrated. Both sources are normally useful in social domains (hence the high hyperprior), but both have a reliability of 0.5 in our experiment. However, although the influence of the former is (eventually) correctly weighted down, at least when participants are explicitly guessing the agent's goal, the same is not true for the latter. This difference could be explained by considering that our participants might have "explicit" access only to the gaze, making it easier to be modulated compared with the hand preshape, which might be processed more automatically.

In sum, our results suggest that gaze information can affect the ability to predict the outcome of others' actions, but only when no other information about their behavioral intention is provided. Conversely, our experiments provide evidence that when motor cues such as a preshaped hand with a given grip are available and might help in selecting action targets, people automatically tend to capitalize on such motor cues despite their unreliability, thus turning out to be much more accurate and fast in predicting the goal of the observed action, even when contrasting sources of evidence such as the actor's gaze direction are presented simultaneously.

#### GRANTS

G. Pezzulo was supported by European Union Seventh Framework Program Grant FP7-ICT-270108 (Goal-Leaders).

#### DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

#### AUTHOR CONTRIBUTIONS

E.A., G.P., and M.C. conception and design of research; E.A. performed experiments; E.A. analyzed data; E.A., G.P., and M.C. interpreted results of experiments; E.A. prepared figures; E.A., G.P., and M.C. drafted manuscript; E.A., G.P., and M.C. edited and revised manuscript; E.A., G.P., and M.C. approved final version of manuscript.

#### REFERENCES

Ambrosini E, Costantini M, Sinigaglia C. Grasping with the eyes. *J Neurophysiol* 106: 1437–1442, 2011.

<sup>2</sup> It is interesting to note that an experimental manipulation similar to that adopted here in terms of congruence between the hand preshape and the size of the target object has been previously applied in an action execution study (Begliomini et al. 2007b). The neural and kinematic results show that the mismatch between the grasp type and the target object size affected both the agent's action kinematic and the cortical activation in his/her visuomotor grasping network. One could argue that these effects may have affected our preshape congruence results, given the strict link between action execution and observation processes. However, it should be noted that a more recent action observation study using a similar experimental manipulation (Cavallo et al. 2011) failed to find any effect of the congruence between the observed grasp and the agent's target object, and thus further investigations are needed to resolve this issue.



- Ambrosini E, Reddy V, de Looper A, Costantini M, Lopez B, Sinigaglia C. Looking ahead: anticipatory gaze and motor ability in infancy. *PLoS One* 8: e67916, 2013.
- Ambrosini E, Sinigaglia C, Costantini M. Tie my hands, tie my eyes. *J Exp Psychol Hum Percept Perform* 38: 263–266, 2012.
- Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang* 59: 390–412, 2008.
- Bach P, Bayliss A, Tipper S. The predictive mirror: interactions of mirror and affordance processes during action observation. *Psychon Bull Rev* 18: 171, 2011.
- Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I. The “Reading the Mind in the Eyes” Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J Child Psychol Psychiatry* 42: 241–251, 2001.
- Bates DM, Maechler M, Bolker B. *lme4: Linear mixed-effects models using Eigen and Eigen++* (Online). R package version 0.999999-0, <http://CRAN.R-project.org/package=lme4> [2012].
- Becchio C, Bertone C, Castiello U. How the gaze of others influences object processing. *Trends Cogn Sci* 12: 254–258, 2008.
- Becchio C, Cavallo A, Begliomini C, Sartori L, Feltrin G, Castiello U. Social grasping: from mirroring to mentalizing. *Neuroimage* 61: 240–248, 2012.
- Begliomini C, Wall MB, Smith AT, Castiello U. Differential cortical activity for precision and whole-hand visually guided grasping in humans. *Eur J Neurosci* 25: 1245–1252, 2007a.
- Begliomini C, Caria A, Grodd W, Castiello U. Comparing natural and constrained movements: new insights into the visuomotor control of grasping. *PLoS One* 2: e1108, 2007b.
- Castiello U. Understanding other people’s actions: intention and attention. *J Exp Psychol Hum Percept Perform* 29: 416–430, 2003.
- Cavallo A, Sartori L, Castiello U. Corticospinal excitability modulation to hand muscles during the observation of appropriate versus inappropriate actions. *Cogn Neurosci* 2: 83–90, 2011.
- Costantini M, Ambrosini E, Sinigaglia C. Does how I look at what you’re doing depend on what I’m doing? *Acta Psychol (Amst)* 141: 199–204, 2012a.
- Costantini M, Ambrosini E, Sinigaglia C. Out of your hand’s reach, out of my eyes’ reach. *Q J Exp Psychol (Hove)* 65: 848–855, 2012b.
- Costantini M, Ambrosini E, Cardellicchio P, Sinigaglia C. How your hand drives my eyes. *Soc Cogn Affect Neurosci* 9: 705–711, 2014.
- Dindo H, Zambuto D, Pezzulo G. Motor simulation via coupled internal models using sequential Monte Carlo. *IJCAI (US)* 22: 2113–2119, 2011.
- Ehrsson HH, Fagergren A, Jonsson T, Westling G, Johansson RS, Forssberg H. Cortical activity in precision-versus power-grip tasks: an fMRI study. *J Neurophysiol* 83: 528–536, 2000.
- Ernst MO, Bulthoff HH. Merging the senses into a robust percept. *Trends Cogn Sci* 8: 162–169, 2004.
- Falck-Ytter T, Gredeback G, von Hofsten C. Infants predict other people’s action goals. *Nat Neurosci* 9: 878–879, 2006.
- Flanagan JR, Johansson RS. Action plans used in action observation. *Nature* 424: 769–771, 2003.
- Freeman JB, Ambady N. MouseTracker: software for studying real-time mental processing using a computer mouse-tracking method. *Behav Res Methods* 42: 226–241, 2010.
- Friesen CK, Kingstone A. The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychon Bull Rev* 5: 490–495, 1998.
- Frith K. The free-energy principle: a unified brain theory? *Nat Rev Neurosci* 11: 127–138, 2010.
- Friston K, Mattout J, Kilner J. Action understanding and active inference. *Biol Cybern* 104: 137–160, 2011.
- Frith CD. The social brain? *Philos Trans R Soc Lond B Biol Sci* 362: 671–678, 2007.
- Galati G, Committeri G, Spitoni G, Aprile T, Di Russo F, Pitzalis S, Pizzamiglio L. A selective representation of the meaning of actions in the auditory mirror system. *Neuroimage* 40: 1274–1286, 2008.
- Gentilucci M, Castiello U, Corradini ML, Scarpa M, Umiltà C, Rizzolatti G. Influence of different types of grasping on the transport component of prehension movements. *Neuropsychologia* 29: 361–378, 1991.
- Gredeback G, Johnson S, von Hofsten C. Eye tracking in infancy research. *Dev Neuropsychol* 35: 1–19, 2010.
- Gredeback G, Melinder A. Infants’ understanding of everyday social interactions: a dual process account. *Cognition* 114: 197–206, 2010.
- Gredeback G, Stasiewicz D, Falck-Ytter T, Rosander K, von Hofsten C. Action type and goal type modulate goal-directed gaze shifts in 14-month-old infants. *Dev Psychol* 45: 1190–1194, 2009.
- Hudson M, Jellema T. Resolving ambiguous behavioral intentions by means of involuntary prioritization of gaze processing. *Emotion* 11: 681–686, 2011.
- Jeannerod M. *The Neural and Behavioural Organization of Goal-Directed Movements*. Oxford: Oxford University Press, 1988.
- Kanakogi Y, Itakura S. Developmental correspondence between action prediction and motor ability in early infancy. *Nat Commun* 2: 341, 2011.
- Kilner JM, Friston KJ, Frith CD. Predictive coding: an account of the mirror neuron system. *Cogn Process* 8: 159–166, 2007.
- Kording KP, Wolpert DM. Bayesian decision theory in sensorimotor control. *Trends Cogn Sci* 10: 319–326, 2006.
- Manera V, Becchio C, Cavallo A, Sartori L, Castiello U. Cooperation or competition? Discriminating between social intentions by observing prehensile movements. *Exp Brain Res* 211: 547–556, 2011.
- Montefinese M, Ambrosini E, Fairfield B, Mammarella N. The “subjective” pupil old/new effect: is the truth plain to see? *Int J Psychophysiol* 89: 48–56, 2013.
- Montefinese M, Ambrosini E, Fairfield B, Mammarella N. Semantic significance: a new measure of feature salience. *Mem Cognit* 42: 355–369, 2014.
- Nummenmaa L, Calder AJ. Neural mechanisms of social attention. *Trends Cogn Sci* 13: 135–143, 2009.
- O’Reilly JX, Schuffelgen U, Cuell SF, Behrens TE, Mars RB, Rushworth MF. Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proc Natl Acad Sci USA* 110: E3660–E3669, 2013.
- Pezzulo G. Studying mirror mechanisms within generative and predictive architectures for joint action. *Cortex* 49: 2968–2969, 2013.
- Pezzulo G, Donnarumma F, Dindo H. Human sensorimotor communication: a theory of signaling in online social interactions. *PLoS One* 8: e79876, 2013.
- Pierro AC, Becchio C, Tubaldi F, Turella L, Castiello U. Motor ontology in representing gaze-object relations. *Neurosci Lett* 430: 246–251, 2008.
- Preusschoff K, ‘t Hart BM, Einhäuser W. Pupil dilation signals surprise: evidence for noradrenaline’s role in decision making. *Front Neurosci* 5: 115, 2011.
- Quené H, van den Bergh H. Examples of mixed-effects modeling with crossed random effects and with binomial data. *J Mem Lang* 59: 413–425, 2008.
- Ramsey R, Cross ES, Hamilton AF. Eye can see what you want: posterior intraparietal sulcus encodes the object of an actor’s gaze. *J Cogn Neurosci* 23: 3400–3409, 2011.
- Ricciardelli P, Iani C, Lugli L, Pellicano A, Nicoletti R. Gaze direction and facial expressions exert combined but different effects on attentional resources. *Cogn Emot* 26: 1134–1142, 2012.
- Rizzolatti G, Sinigaglia C. The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nat Rev Neurosci* 11: 264–274, 2010.
- Rotman G, Troje NF, Johansson RS, Flanagan JR. Eye movements when observing predictable and unpredictable actions. *J Neurophysiol* 96: 1358–1369, 2006.
- Salvucci DD, Goldberg JH. Identifying fixations and saccades in eye-tracking protocols. In: *ETRA 2000-Proceedings of the Eye Tracking Research and Applications Symposium, Palm Beach Gardens, FL, November 6–8, 2000*, p. 71–78.
- Sartori L, Becchio C, Castiello U. Cues to intention: the role of movement information. *Cognition* 119: 242–252, 2011.
- Tomeo E, Cesari P, Aglioti SM, Urgesi C. Fooling the kickers but not the goalkeepers: behavioral and neurophysiological correlates of fake action detection in soccer. *Cereb Cortex* 23: 2765–2778, 2013.