Grasping with the eyes

Ettore Ambrosini,1* Marcello Costantini,1* and Corrado Sinigaglia2

1Laboratory of Neuropsychology and Cognitive Neuroscience, Department of Neuroscience and Imaging, University “G. d’Annunzio,” and Institute for Advanced Biomedical Technologies, Foundation University “G. d’Annunzio,” Chieti; and 2Department of Philosophy, University of Milan, Milan, Italy

Submitted 10 February 2011; accepted in final form 3 June 2011

Ambrosini E, Costantini M, Sinigaglia C. Grasping with the eyes. J Neurophysiol 106: 1437–1442, 2011. First published June 8, 2011; doi:10.1152/jn.00118.2011.—When observing someone else acting on an object, people implement goal-specific eye movement programs that are driven by their own motor representation of the observed action. Usually, however, we observe people acting in contexts where more objects, different in shape and size, are present. Is our brain able to select the intended target even when there are different objects in the visual scene? And if this is the case, what kind of information does our motor system capitalize on? 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. In a control condition, the actor merely reached for and touched one of the two objects without preshaping her hand according to the target features. Results showed higher accuracy and earlier saccadic movements when participants observed an actually grasping hand than when they observed a mere reaching hand devoid of any kind of target-related preshaping. This clearly suggests that the hand preshaping provided the observer with enough motor cues to proactively and reliably saccade toward the object to be grasped, thus identifying it even when the action target was not previously known. Our findings strongly corroborate the direct matching hypothesis suggesting that in processing others’ actions, we take advantage of the same motor knowledge that enables us to efficiently perform those actions.

MANIPULATING OBJECTS requires specific goal-related saccadic eye movements (Hayhoe and Ballard 2005; Land 2006, 2009). These eye movements have been demonstrated to be proactive in nature, seeking out the information needed by the motor system in planning and monitoring the execution of a given action (Johansson et al. 2001). In grasping actions, for instance, eyes typically shift toward objects that will be eventually picked up, while being rarely grabbed by action irrelevant objects (Rothkopf et al. 2007).

In a seminal study, Flanagan and Johansson (2003) showed that when people observe object-related manual actions (e.g., block-stacking actions), the coordination between their gaze and the actor’s hand is highly similar to the gaze-hand coordination when they perform those actions themselves. In both cases, people proactively shift their gaze to the target sites, thus anticipating the outcome of the actions without attending to their visual unfolding.

According to the authors, the proactivity of eye movements in both action execution and action observation supports the direct matching hypothesis (Flanagan and Johansson 2003). This hypothesis postulates that observing actions performed by others elicits a motor activation in the brain of the observer similar to that which occurs when she plans her own actions (Rizzolatti et al. 2001; Rizzolatti and Sinigaglia 2010). Consequently, when observing someone else acting on an object, people implement goal-specific eye movement programs that are driven by their own motor representation of the observed action.

Flanagan and Johansson (2003) employed a paradigm in which the outcome of the observed actions was fully predictable, being the target objects previously known and the sequence of grasping and stacking movements always the same. Usually, however, we observe people acting in contexts where more objects are present. In particular, we witness other people manipulating one object among many others that differ in shape, size, and/or texture. Thus the question arises as to whether eye movements are proactive even in these cases, as well as whether such proactivity could be accounted for in terms of the direct matching hypothesis. In other words, could our motor representations of the observed actions proactively drive our gaze toward their target objects even when the latter are not known in advance? And if this is the case, what kind of information does our motor system capitalize on?

To tackle these issues we used a task in which participants simply observed an actor reaching for and grasping one of two objects while eye movements were recorded. Notably, the two objects differed as to the kind of grip required to pick them up, namely, precision grip or whole hand prehension. In a control condition, the actor merely reached for and touched one of the two objects without grasping it. It is well known that the finger configuration proper to a given grip is but the end result of a motor sequence beginning well ahead of the contact with the object, being the hand preshaped according to the intrinsic features of the latter during its movement toward it (Jeannerod et al. 1995). Thus, if the direct matching hypothesis holds, we expect that when observing grasping actions, people will capitalize on the same motor information about the hand preshaping that is crucial for planning and monitoring their own grasping actions. In particular, we expect that although observing a moving hand might per se be enough to induce proactive saccadic movements, only the observation of the critical phase of the hand shaping process will allow one to reliably identify the intended target, gazing at it well before the approaching hand. We therefore predict that saccadic eye movements will be both more accurate and faster in grabbing the intended target when observing a hand grasping objects by preshaping a given grip than when observing the hand merely reaching for
and touching them, without conveying any motor cue about the grip.

METHODS

Participants. Fifteen subjects (mean age = 25 yr, SD = 2.6 yr) participated in the experiment. All participants provided informed consent. The subjects were all right-handed, did not require corrective lenses, and had no history of ophthalmological or neurological disease. The procedures were approved by the Ethical Committee of the “G. d’Annunzio” University, Chieti, Italy, and were in accordance with the ethical standards of the 1964 Declaration of Helsinki.

Apparatus and stimuli. We recorded participants’ eye movements while they observed the stimulus videos. Subjects were seated in a comfortable chair in front of a 17-in. computer monitor that was positioned on the frontoparallel plane at a distance of 57 cm. Their heads were stabilized by means of a chin and a head rest. An infrared video-based eye-tracking device (RK-826PCI pupil/corneal tracking heads were stabilized by means of a chin and a head rest. An infrared positioned on the frontoparallel plane at a distance of 57 cm. Their comfortable chair in front of a 17-in. computer monitor that was while they observed the stimulus videos. Subjects were seated in a movement toward either a small or a large tomato (targets), both from the side view an actor performing an unpredictable reaching/movement, depending on the target) was clearly visible as /H11003 pixels, corresponding at 21.25°

The experimental videos (AVI format, 30 frames/s, 640 × 480 pixels, corresponding at 21.25° × 16.88° of visual angle) showed from the side view an actor performing an unpredictable reaching movement toward either a small or a large tomato (targets), both located on a table at a distance of ~70 cm from the actor’s torso and 10 cm apart from each other. To counterbalance the hand trajectories, we used four different target layouts (see Fig. 1). Only the actor’s right arm and torso were visible in the videos, and all the arm movements started with the actor’s hand resting on the table immediately in front of his torso. In half of the videos, the actor performs a reach-to-touch movement with the closed fist (no shape condition), whereas in the other half, the actor performs a reach-to-grasp movement during which the prehension of the hand (either a precision or a whole hand prehension, depending on the target) was clearly visible as soon as the movement started (preshape condition) (see Fig. 2). Thus there were a total of 16 different stimulus videos (4 target layouts × 2 movement types × 2 targets). All the experimental videos lasted 2,500 ms: the first 1,000 ms depicted the actor’s hand resting on the table in the starting position with a white fixation cross superimposed on it (fixation phase); the following 1,000 ms showed the entire arm movement, i.e., from the earliest detectable movement of the hand to the hand-object contact (movement phase); and the last 500 ms consisted of the last frame of the stimulus video that was shown as still (contact phase).

The presentation of the stimuli and the recording of the participants’ responses were controlled by custom software (developed by Gaspare Galati at the Department of Psychology, Sapienza Università di Roma, 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 at the Laboratory of Neurobiology, Wellcome Department of Imaging Neuroscience, University College London).

Procedure. At the beginning of each session, gaze position was calibrated using a standard nine-point calibration procedure. Subjects were asked to fixate each of the calibration points distributed over a 3 × 3 grid that covered the entire screen uniformly. Each sampled data point during the experiment was calibrated off-line using separate multiple linear regressions in the horizontal (x) and vertical (y) dimensions. The terms included in the regression were x, y, x^2, y^2, xy, x’y’, xy’, and x’y^2. The resultant regression coefficients were used to scale the data points obtained in the registration session following the calibration measurements. In this way, the gaze trace was represented in pixel units, thus permitting its overlap on the corresponding stimulus video.

Every trial began with the presentation of a white cross at the center of the screen that participants were required to fixate. After 2,500 ms, the stimulus video was presented, and the participants were asked to move their gaze on the fixation cross over the actor’s hand (resting in the starting position) until its disappearance and then to simply watch the video. There were four recording sessions, each composed of 20 trials, for a total of 80 trials (5 repetitions for each of the 16 stimulus videos). In each session, the stimuli were balanced for both movement type and target. Both the order of trials within a session and the order of sessions were randomized.

Data analysis. Data analysis was performed on the 1,000 ms of the gaze trace corresponding to the entire actor’s movement phase using an I-VT (Velocity-Threshold Identification) algorithm that automatically detected saccades and fixations by means of an eye velocity threshold. This algorithm was taken from Salvucci and Goldberg (2000) and modified by adding a temporal criterion (Erkelens and Vogels 1995) and a double velocity threshold to mitigate the instrument noise, permit the identification of smooth pursuits, and prevent...

Fig. 1. Four different layouts used in the experiment.
For each trial, we created three areas of interest (AOI), covering the hand (hand AOI), the intended target (target AOI), and the nonintended target (no-target AOI), respectively. Each AOI was actually 0.07° larger than the real stimulus to compensate for noise in the eye-tracking system. Hand AOI was a dynamic offline AOI; i.e., it was manually added frame by frame in order to match gaze trace with the moving hand after the recording (Papenmeier and Huff 2010). For each trial, we calculated the number of saccades, the time at which the first saccade started (onset time), and the arrival time of the last saccade on the target AOI (arrival time).

Given that the actor performed the movements as he would in real life, there were slight variations in movement kinematics with respect to the intended target. Indeed, differences do exist in the reaching phase (especially in its deceleration phase) of reach-to-grasp actions directed to objects of different sizes. It is well known that a grasping hand directed to small objects reaches the peak velocity earlier and has a longer deceleration phase compared with a grasping hand directed to large objects (Gentilucci et al. 1991).

There were a total of 1,200 recorded trials (80 trials × 15 participants). Of these, five trials were discarded because the data were corrupted. To assess experimental effects, we used repeated-measures ANOVAs with shape (no shape vs. preshape) and target (small vs. large target) as within-subject factors. There were four experimental conditions, corresponding to the four types of observed hand actions, namely, no shape–large target, no shape–small target, preshape–large target, and preshape–small target. Post hoc Tukey’s honestly significant difference test was used when necessary.

**RESULTS**

For this reason, and to control for residual differences in movement kinematics, we chose not to use a fixed time threshold to calculate the gaze onset and arrival times. Instead, for each trial, onset and arrival times were calculated relative to the hand movement onset and offset, defined using a velocity threshold (10°/s, roughly corresponding to 0.5 m/s or 3 pixels/frame). It should be noted that such a criterion is more conservative compared with hand-target contact. To calculate the hand velocity, we first determined from each frame of each video the location of the radial styloid process of the wrist; next, we interpolated at 120 Hz the trajectory of this marker and then calculated its point-to-point velocity.
the number of saccades. It should be reminded that the movement phase lasted 1,000 ms, corresponding to the entire actor’s movement. We did not take into consideration the trials in which participants either fixated or pursued the hand AOI (5.9 and 7.5% of the recorded trials, respectively) or performed a saccade toward the nonintended target (no-target AOI, 11.6% of the recorded trials).

The percentage of correct trials was entered in a repeated-measures ANOVA as described above. The main factor shape was significant \( F(1,14) = 38.9, P < 0.001 \) with higher accuracy on preshape trials (82%) compared with no-shape trials (65%). The main factor target was also significant \( F(1,14) = 92.3, P < 0.001 \) with higher accuracy on large-target (86%) compared with small-target trials (61%). The shape \( \times \) target interaction was also significant \( F(1,14) = 48.8, P < 0.001 \); see Fig. 3. The interaction was explained by higher impact of preshape on small target (no shape 48%; preshape 74%; \( P < 0.01 \)) compared with large-target trials (no shape 82%; preshape 91%; \( P < 0.01 \)).

Onset and arrival time. Onset and arrival times of saccades made in correct trials were calculated from individual gaze traces as described in METHODS. The ANOVA conducted on the onset times revealed a significant main effect of shape \( F(1,14) = 19.6, P < 0.001 \), with earlier onset time on preshape (193 ms) than no-shape trials (221 ms). The main effect of target was also not significant (\( F \) values < 1).

Arrival times analysis revealed a significant main effect of shape \( F(1,14) = 235.4, P < 0.001 \), with earlier arrival times on preshape (-143 ms) than no-shape trials (10 ms). The main effect of target was also significant \( F(1,14) = 71.9, P < 0.001 \), with earlier arrival times on large-target (-123 ms) compared with small-target trials (-9 ms). The shape \( \times \) target interaction was also significant \( F(1,14) = 33.1, P < 0.001 \); see Figs. 4 and 5. The interaction was explained by higher impact of preshape on small-target (no shape = 90 ms; preshape = -109 ms; \( P < 0.01 \)) compared with large-target trials (no shape = -70 ms; preshape = -176 ms; \( P < 0.01 \)).

Number of saccades. The mean number of saccades performed in each condition was entered in an ANOVA as described above, which revealed the significant main effect of shape \( F(1,14) = 184.8, P < 0.001 \), with fewer saccades made during preshape (1.62) than no-shape trials (2.06). The main factor target was also significant \( F(1,14) = 33.9, P < 0.001 \), with fewer saccades made in shifting gaze toward the large target (1.68) compared with the small target (2.01). Moreover, the shape \( \times \) target interaction turned out to be significant \( F(1,14) = 10.6, P < 0.01 \); post hoc analysis revealed that the visual information provided by the hand preshape had a higher impact on small-target (no shape = 2.29; preshape = 1.73; \( P < 0.001 \)) compared with large-target trials (no shape = 1.83; preshape = 1.52; \( P < 0.001 \)).

DISCUSSION

In this study 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. There were two main findings.

First, the eye movement accuracy was significantly higher when participants observed an actually grasping hand than when they observed a mere touching hand devoid of any target-related preshaping. Interestingly, our results showed that the number of trials in which participants successfully proactively gazed at the target was significantly higher when participants witnessed the preshaping of the actor’s hand than when they witnessed the actor’s unshaped hand merely approaching the object. Note that in the no-shape condition, when the observed hand was approaching the large target, participants saccaded toward the intended object in a percentage of trials above chance; this was not the case when they observed the actor’s hand moving toward the small target. A possible accounting for such different gaze behavior is that participants selected the large object as the default target, this object being more salient than the small one.

Second, as for the proactive saccadic time course, we found not only that the onset time of the first saccadic movement was earlier in the preshape than in the no-shape condition (~30 ms) but also and above all that gaze proactively reached the object to be grasped 153 ms earlier in the preshape than in the no-shape condition. It is worth noting that the observation of the preshaping hand impacted more on the saccadic arrival time when the target object was small, thus requiring a precision grip, than when it was large, thus requiring a whole hand prehension. This could be accounted for by the motor information to be processed, given that that related to the precision grip is more detailed and time-consuming (Gentilucci et al. 1991). The situation was different in the no-shape condition, where participants proactively gazed toward the large target object only. Indeed, observing the unshaped hand moving toward small objects did not proactively drive eye movements, because the observer gazed toward the target 90 ms later than the time at which the actor’s hand reached it. As with the case of accuracy, the different arrival times in the no-shape condition are very likely due to the implementation of a default strategy for selecting the intended object. The comparison of the number of saccades indicates that, in the absence of any motor cues, participants essentially gazed toward the large object, shifting their gaze to the small one when they realized that the latter was the target.

Previous evidence for the proactivity of eye movements during the observation of others’ actions has been provided by Flanagan and Johansson (2003). They showed that when people observe a block-stacking task, the coordination between their gaze and the actor’s hand is proactive rather than reactive, thus being highly similar to the gaze-hand coordination when they perform the task themselves. According to the authors, this evidence strongly supports the direct matching hypothesis, postulating that both executing a given motor behavior and witnessing such a motor behavior performed by some else recruit the same motor resources (Rizzolatti et al. 2001; Rizzolatti and Sinigaglia 2010). This hypothesis has been further corroborated by Falck-Ytter et al. (2006). They demonstrated not only that proactive goal-directed eye movements in adults result from the direct matching of an observed action with the motor representation of that action but also and above all that infants proactively gaze toward the target object of others’ actions at the same age as they become able to perform those actions themselves. More recently, Gredebäck et al. (2009) showed that observing different actions performed with different motor intentions modulates the latency of proactive gaze shifts in infants.

Our results extend the findings of Flanagan and Johansson (2003) by assessing the impact of the motor components of the observed action, such as hand preshaping, on the proactivity of the observer’s eye movements. The magnitude of the gaze proactivity in our study is comparable with that found by Flanagan and Johansson (2003), turning out to be smaller than that found in other studies (e.g., Falck-Ytter et al. 2006; Gredebäck et al. 2009). This is likely to be due to the different kinds of actions involved in the tasks (e.g., grasping vs. placing actions) and therefore to the different kinds of motor cues needed to perform those tasks.

It is worth noting that Flanagan and Johansson (2003) employed a paradigm in which the sequence of the observed stacking movements was fully predictable, with the task to stack the blocks from the widest to the narrowest always being the same. Although this paradigm was successful in comparing eye movements while observing and executing actions, it did not help in highlighting the gaze behavior when the action targets were not previously known. However, most actions we witness are usually related to one object among others, different in shape, size, and/or texture, and very often we do not know in advance which one it will be.

To successfully manipulate objects, an agent needs not only to proactively look at them but also to transform the objectual features into a suitable finger configuration, thus preshaping her own hand a long time before approaching the corresponding targets (Jeannerod 1988; Jeannerod et al. 1995). Very recently Bach et al. (2011) demonstrated that this is also the case for observed actions. Indeed, they showed that action observation not only relies on a matching process but implies the automatic detection of the potential for successful action, based on an integration of object and action features.

The present study extends these findings, demonstrating for the first time the critical role played by the processing of hand preshaping in driving the observer’s proactive eye movements to the action target. Our data clearly show that the hand preshaping provided the observer with enough motor cues to proactively and reliably saccade toward the object to be grasped, thus identifying it even when the action target is not previously known.

These results are also entirely consistent with evidence provided by Rotman et al. (2006). The aim of their study was to compare observers’ eye movements during predictable and unpredictable block-stacking tasks. Participants watched an actor reach for, lift, and replace first one block and then a second one located either closer to or further away from them. They could infer which block would be targeted second but did not have advance information about which block would be reached for and lifted first. The results showed that in all conditions observers engaged gaze in a proactive fashion by looking at the target object ahead of the actor’s hand. Interestingly, during the observation of the unpredictable movements, observers seemed to adopt a default strategy, first shifting their gaze from the start block to the near block and then onto the far block as soon as they assessed that the latter, but not the former, was the target of the actor’s hand.
Taken together, the present data and the findings of Rotman et al. (2006) indicate that someone else’s actions might proactively drive the observer’s eye movements even when the targets of those actions are not previously known, suggesting that observers may implement different strategies according to the information available to them. Indeed, when the available motor cues cannot be helpful in immediately identifying the objects targeted by others’ actions, people tend to adopt a default strategy, saccading first toward the most salient stimulus and then shifting to the other one as soon as they realize that the latter is the target. On the contrary, 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 information, thus turning out to be much more accurate and fast in gazing at the object to be manipulated by the other’s hand.

Overall, this is in line with what is predicted by the direct matching hypothesis. Indeed, this hypothesis postulates not only that in processing others’ actions, we take advantage of the same motor knowledge that enables us to efficiently perform those actions, but also that such a processing might be modulated by our own motor repertoire as well as by the available visuomotor information (Rizzolatti and Sinigaglia 2010). This does not rule out the possibility that top-down processing of others’ actions based on contextual information might help one to identify the target objects, as measured by the object to be manipulated by the other’s hand.

Overall, this is in line with what is predicted by the direct matching hypothesis. Indeed, this hypothesis postulates not only that in processing others’ actions, we take advantage of the same motor knowledge that enables us to efficiently perform those actions, but also that such a processing might be modulated by our own motor repertoire as well as by the available visuomotor information (Rizzolatti and Sinigaglia 2010). This does not rule out the possibility that top-down processing of others’ actions based on contextual information might help one to identify the target objects, as measured by the object to be manipulated by the other’s hand.

GRANTS

C. Sinigaglia was supported by a grant from the San Paolo Foundation. M. Costantini and C. Sinigaglia were supported by the EU Grant TESIS.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

REFERENCES


