What first drives visual attention during the recognition of object-directed actions? The role of kinematics and goal information

2 authors:

Jérémy Decroix
Université de Lille
6 PUBLICATIONS   **25** CITATIONS

SEE PROFILE

Solène Kalénine
Université de Lille
46 PUBLICATIONS   **941** CITATIONS

SEE PROFILE

1   **Title:** What first drives visual attention during the recognition of object-directed actions? The

2   role of kinematics and goal information.

3   **Author names:** Jérémy Decroix & Solène Kalénine

4   **Affiliations :** Univ. Lille, CNRS, CHU Lille, UMR 9193 - SCALab - Sciences Cognitives et

5   Sciences Affectives, F-59000 Lille, France

6   **Authors' addresse:**

7   Laboratoire SCALab UMR CNRS 9193

8   Université de Lille, SHS

9   Batiment A2 - Niveau forum +2

10   Domaine universitaire du Pont de Bois, BP 149

11   59653 Villeneuve d'Ascq Cedex, France

12

13   **Corresponding author:** Solène Kalénine

14   Mail: solene.kalenine@univ-lille.fr

15   Phone: +33 3 20 41 73 25

16

Abstract: The recognition of others' object-directed actions is known to involve the decoding of both the visual kinematics of the action and the action goal. Yet whether action recognition is first guided by the processing of visual kinematics or by a prediction about the goal of the actor remains debated. In order to provide experimental evidence to this issue, the present study aimed at investigating whether visual attention would be preferentially captured by visual kinematics or by action goal information when processing others' actions. In a visual search task, participants were asked to find correct actions (e.g. drinking from glass) among distractor actions. Distractors actions contained grip and/or goal violations and could therefore share the correct goal and/or the correct grip with the target. The time course of fixation proportion on each distractor action has been taken as an indicator of visual attention allocation. Results show that visual attention is first captured by the distractor action with similar goal. Then the withdrawal of visual attention from the action distractor with similar goal suggests a later attentional capture by the action distractor with similar grip. Overall, results are in line with predictive approaches of action understanding, which assume that observers first make a prediction about the actor's goal before verifying this prediction using the visual kinematics of the action.

Key-words: visual attention, predictive approaches, action semantics, object-directed action, eyetracking

Word count : 217

1 Actions are complex cognitive phenomena and can be described at different levels of

2 abstraction, from abstract action intentions to the description of the mechanistic properties of

3 movements (Jacob & Jeannerod, 2005; Kilner, 2011; Urgesi, Candidi, & Avenanti, 2014).

4 Decades of research on action planning have highlighted the hierarchical structure of actions,

5 where higher-goals lead to the selection of sub-goals that are then translated into appropriate

6 motor programs (Cooper, Ruh, & Mareschal, 2014; Thill, Caligiore, Borghi, Ziemke, &

7 Baldassarre, 2013; van Elk, van Schie, & Bekkering, 2014; Wolpert, Doya, & Kawato, 2003).

8 Actions are thus organized and goal-directed movements.

9 Actions are not only planned around goals, but they are also perceived as goal-directed. Thus,

10 humans most likely identify their own actions and the actions of others as goal-directed (Hrkać,

11 Wurm, & Schubotz, 2014; Novack, Wakefield, & Goldin-Meadow, 2016; Vallacher & Wegner,

12 1987, 2012; Zacks, Tversky, & Iyer, 2001). Many pieces of evidence in this direction can be

13 found in the visual attention literature. During the observation of reach and grasp movements,

14 both children and adults make proactive gaze movements towards the expected landing point

15 of the action (Ambrosini, Costantini, & Sinigaglia, 2011; Flanagan & Johansson, 2003;

16 Flanagan, Rotman, Reichelt, & Johansson, 2013; Geangu, Senna, Croci, & Turati, 2015), which

17 suggests that observers do not simply follow the movement course as it unravels but predict

18 and anticipate the goal of the action. This bias of interpreting actions as goal-directed seems to

19 arise quite early in development. Infants are indeed able to track others' goals (Buresh &

20 Woodward, 2007) and they show a renewal of attention when an actress stops her movement

21 without achieving her goal (Baldwin, Baird, Saylor, & Clark, 2001). Overall, these data

22 highlight the importance of goals during the perception and the recognition of actions

23 performed by others (Ocampo & Kritikos, 2011).

24 Yet the decoding of action goals may be less straightforward during the perception of others'

25 actions than during action planning. Indeed, the actor's goal is not readily available for the

observer, and different approaches have been proposed to explain how one succeeds to understand others' goals. Thus, sensorimotor approaches of action understanding have suggested that goals "become *"visible"* in the surface flow of agents' motions" (Ansuini, Cavallo, Bertone, & Becchio, 2014 p.1), that is, the actor's goal emerges in the mind of the observer from the processing of the visual kinematics (i.e., reach trajectory, grip configuration or means; cf. Grafton & Hamilton, 2007) of his/her actions. Several pieces of evidence suggest that this ability may arise in development through repetitive association between movements and their perceptual consequences (Coello & Delevoye-Turrell, 2007; Hunnius & Bekkering, 2014). Accordingly, proactive gaze movements have been shown to be facilitated in the presence of information about the shape of the hand (Ambrosini et al., 2011) or when infants possess sufficient motor expertise about a given action (Ambrosini et al., 2013; Geangu et al., 2015). Attending the motor component of the action must then be required to understand the goal of an actor.

Contrasting with this approach, some authors have highlighted the fact that decoding the goal of the actor on the sole basis of the observed kinematics was possible only in rare cases of unambiguous actions (Hunnius & Bekkering, 2014; Jacob & Jeannerod, 2005). Alternatively, it has been proposed that kinematic processing would be guided by the prior activation of a prediction of the actor's goal driven by contextual information (Kilner, Friston, & Frith, 2007) or by non-motor components of the action, such as the object-tool (Bach, Nicholson, & Hudson, 2014). We will subsequently refer to these approaches as "predictive approaches". Accordingly, it has been shown that prior knowledge about the actor's goal modifies the subsequent processing of the kinematics by the observer (Hudson, Nicholson, Ellis, & Bach, 2016; Hudson, Nicholson, Simpson, Ellis, & Bach, 2016). Similarly, it has been demonstrated that fMRI brain activity during the processing of goals is more similar to the processing of object-tools than to the processing of kinematics (Nicholson, Roser, & Bach, 2017). Goal processing may then rely

more on object-tool information than on information related to visual kinematics. Furthermore, observers attend to information about trajectory to a greater extent for dropping than for placing actions, suggesting that they differentially use kinematic information depending on the action goal of the actor (Loucks & Pechey, 2016).

Together, the available evidence is not conclusive about whether action decoding is first driven by the processing of the visual kinematics (e.g., the grip), or by the processing of non-motor components (e.g., the object-tool) of the action, and to what extent action decoding is sensitive to the availability of non-motor information about the action. Indeed, studies supporting sensorimotor approaches tend to carry out experiments in which kinematics are the sole discriminant information about the action, whereas predictive approaches tend to present contextual information before the presentation of a target action. Consequently, although we know that both kinematics and goal prediction are involved in action processing, whether action processing is driven by the early decoding of visual kinematics or by a prediction about the actor's goal through the processing of non-motor action components is still unclear. The overall weight of goals in action decisions is not informative either, as visual kinematics may still be processed first (see for example Kilner & Frith, 2008; Tidoni & Candidi, 2016). The spontaneous orientation of visual attention towards visual kinematics or goal-related information may help dissociating the two approaches. Visual attention has indeed been found to impact the processes involved in the decoding of others' actions (D'Innocenzo, Gonzalez, Nowicky, Williams, & Bishop, 2017; Donaldson, Gurvich, Fielding, & Enticott, 2015; Leonetti et al., 2015; Muthukumaraswamy & Singh, 2008; Perry, Troje, & Bentin, 2010; Riach, Holmes, Franklin, & Wright, 2018; Schuch, Bayliss, Klein, & Tipper, 2010; Woodruff & Klein, 2013; Wright et al., 2018) and to be affected by visual kinematics and goal-related information (Humphreys et al., 2013 for review). Yet the temporal dynamics of visual attention allocation on visual kinematics and goal-related information remains to determine.

1 The present study aimed at investigating what captures attention first in an action discrimination

2 task where observers search for correct actions among distractor actions that could have either

3 the same grip or the same goal as the target action. The discrimination task is well-suited to

4 directly and independently oppose grip and goal dimensions. In other words, is visual attention

5 preferentially driven towards grip information or towards non-motor information that may help

6 building a prediction about the goal of the actor?

7 In the present study, we followed the repartition of eye movements during a visual search task

8 to evaluate the influence of grip and goal-related information (e.g. orientation of the object) on

9 the temporal allocation of visuospatial attention. Static photographs of actions were used, which

10 allow displaying both grip and goal information at the exact same time. Grip configuration may

11 not be as predictive of the outcome of the action as the full dynamic kinematics. However,

12 significant changes in grip configuration can still be very informative of whether an action is

13 correct or not overall. Moreover, static visual kinematics have been shown to be particularly

14 important to identify what an actor is doing with an object (Naish, Reader, Houston-Price,

15 Bremner, & Holmes, 2013). Therefore, visual kinematics have been manipulated through

16 changes in grip configuration in our stimuli. Participants were then asked to find a picture

17 displaying a typical object-directed action among distractor action pictures. Distractor pictures

18 displayed either a "similar action goal but a dissimilar grip", or a "similar grip but a dissimilar

19 action goal", or both a "dissimilar action goal and a dissimilar grip". In case observers are first

20 paying attention towards the grip to derive the action goal, "similar grip but dissimilar action

21 goal" distractors should capture visual attention earlier than "similar action goal but dissimilar

22 grip" distractors. Alternatively, if observers first use non-motor information of the action to

23 orient the processing of kinematic information, then "similar action goal but dissimilar grip"

24 distractors should capture visual attention earlier than "similar grip but dissimilar action goal".

25

1   **Methods**

2   **Participants**

3   Twenty-two participants took part in the study[1]. Two participants were left-handed according

4   to the Edinburgh handedness inventory (EHI; Oldfield, 1971) and were then excluded. One

5   participant was excluded because of technical problems during the experimental session.

6   Finally, two participants were excluded because of an atypical pattern of fixation in comparison

7   to the remaining participants (see below). Eighteen participants (mean age 23, age range 18-27,

8   5 males) were then included in the final sample. All were right-handed (mean EHI 96%, from

9   63 to 100%), reported normal or corrected-to-normal vision. They provided written informed

10 consent and were not paid for their participation. The study followed the ethical guidelines of

11 the University of Lille and was in accordance with the declaration of Helsinki (1964, revised in

12 2013).

13   **Stimuli and Design**

14 Twenty objects were selected. For each object, four 512×341 pixel coloured photographs of

15 object-directed actions were designed by crossing the correctness of the grip and goal

16 components of the action: the object-directed action could display a "Correct Grip and Correct

17 Goal", a "Correct Grip only", a "Correct Goal only", or both "Incorrect Grip and Incorrect

18 Goal". Correct grips were defined as the typical grasp-to-use of the object. Incorrect grips then

19 corresponded to an atypical (but not impossible) grasp-to-use of the object. Similarly, goals

20 were considered correct if the typical function of the object could be achieved. Incorrect goals

21 then corresponded to an atypical (but not impossible) goal according to the main function of

---

[1] The sample size of about 20 participants was determined based on previous eye tracking studies using a similar paradigm (e.g. target search in 4-pictures display) in which differences of about 3% of fixation proportion were reported between two pictures in competition (e.g., Kalénine et al. 2012, Cohen *d* of 0.12 for the estimated difference of 3%). A power analysis indicated that twenty participants was sufficient to ensure a statistical power of 0.82 to detect difference in fixation proportion between two pictures (calculated from 50 simulations using the powerCurve function of the R package simr, Green & MacLeod, 2016).

1   the object. Importantly, the incorrect grip did not prevent the correct goal from being achieved.

2   For example, using a power grasp to write with an upright pencil is atypical but does not prevent

3   writing. On the contrary, using a precision grip to write with a pencil upside down does not

4   allow writing although the grip configuration applied on the pencil (the precision grip) is

5   typical. Thus, grip and goal varied independently from one another. An example of the stimuli

6   can be found in Figure 1. The full set of stimuli is available as supplementary materials.

7   [INSERT FIGURE 1 ABOUT HERE]

8   **Procedure**

9   Participants were comfortably seated in front of a 1024 x 768 computer screen in a quiet and

10   darkened room. Head movements were restrained with a chin and forehead rest to reduce

11   measurement errors. The vision was binocular but only the position of the left eye was recorded

12   for all participants. Eye movements were measured continuously with an infrared video-based

13   eye tracking system (EyeLink, SR Research), sampled at 500 Hz. Before each experimental

14   session, the eye tracker was calibrated by asking participants to fixate a set of 9 fixed locations

15   distributed across the screen. After the calibration, instructions were given to each participant

16   and a training session with feedback was provided. The training session included five

17   representative trials with objects that were not in the experimental session. The experimental

18   session was similar to the practice session, but without feedback. Each trial began with a

19   fixation cross in the center on the screen. Participants had to click on the fixation cross to make

20   the display appear. For each reference object, pictures were randomly assigned to the different

21   corners of the screen. The centre of each picture was at 13 degree of visual angle of the centre

22   of the screen. Participants were asked to click on the picture displaying the correct action

23   according to the typical use of the object with the mouse. The "Correct Grip and Correct Goal"

24   picture was defined as the "Target", the "Correct Grip only" picture as the "Grip-distractor",

25   the "Correct Goal only" picture as the "Goal-distractor", and the "Incorrect Grip and Incorrect

Goal" picture as the "Unrelated-distractor". Overall, there were 20 trials corresponding to each reference object. Eye movements were recorded from the beginning of each trial until the mouse-click response on the images.

Fixation proportion

Data analysis followed a procedure previously used in eye tracking studies to capture the evolution of eye movement distribution across time (Kalénine, Mirman, Middleton, & Buxbaum, 2012; Lee, Middleton, Mirman, Kalénine, & Buxbaum, 2013; Lee, Mirman, & Buxbaum, 2014; Mirman, Dixon, & Magnuson, 2008; Mirman & Magnuson, 2009). Four areas of interest (AOI) associated with the displayed pictures were defined as the four 512 x 341 pixel quadrants of the 1024 x 643 pixel computer screen. We considered that participants fixated a given action type ("Target", "Grip-distractor", "Goal-distractor" and "Unrelated-distractor") when their gaze fell into the corresponding AOI. Fixation proportion on each action type was calculated over 50 ms time bins in order to reduce the noise in the fixation estimates and to facilitate statistical model fitting (see Data analysis section). For each time bin of each participant or each item, mean fixation proportion for each action type was computed by dividing the number of fixations on this action type by the total number of trials to avoid the selection bias introduced by varying trial-termination times (cf. Kukona, Fang, Aicher, Chen, & Magnuson, 2011; Mirman & Magnuson, 2009; Mirman, Strauss, Dixon, & Magnuson, 2009).

Saliency Maps

The experiment aimed at assessing which action component first drives visual attention when identifying a target action among distractors. Yet visual selective attention is largely influenced by the visual properties of the image to explore (e.g., colour, spatial orientation, intensity etc.). In order to partial out the effect of possible differences in low-level visual features between the four images on gaze behaviour during target action visual search, saliency maps were computed with the Saliency ToolBox for each stimuli (Walther & Koch, 2006). Saliency values were then

extracted for each pixel and averaged across each area of interest (see "Fixation proportion" section). A saliency index was therefore available for each of the four pictures ("Target", "Grip-distractor", "Goal-distractor", "Unrelated-distractor") of each of the 20 displays. Paired comparisons showed a perceptual advantage for the "Goal-distractor" over the "Grip-distractor" ($t_{19} = -5$, $p < .001$). Saliency indices were thus added as covariate in a complementary by-item analysis.

Data Analysis

The temporal dynamics of fixations on the two "Grip-" and "Goal-distractor" pictures were compared in order to determine whether visual attention is first captured by grip or by goal information[2]. To capture the effect of time, fixation proportions over time were fitted as a function of fourth-order orthogonal polynomials. Orthogonal polynomials are well suited to characterize different behaviours of the fixation curves (see Mirman, 2014 for an introduction to Growth Curve Analysis). Fourth order polynomials were chosen since they have been proven successful to capture the rise and fall of the fixation curves of competing distractors during target identification (Mirman, 2014; Mirman et al., 2008). The intercept reflects differences in the overall height of the curve between conditions. In the present study intercept differences between goal and grip distractors would not inform on which action dimension is processed first and was not of primary interest. Differences in timing between grip and goal processing would be particularly reflected by differences on the linear (1st order) and/or on the cubic (3rd order) time terms (Kalénine et al., 2012; Lee et al., 2013). If visual attention is first captured by grip information, then we should observe earlier fixations on the "Grip-distractor" in comparison to the "Goal-distractor". This would be reflected by a more negative linear estimate (slope) or cubic estimate for the goal compared to the grip fixation curve. Conversely, we

---

[2] As the target fixation curve was incomparable to distractor fixation curves, the target was not included in the analysis (see Figure 2). The unrelated distractor was not added in the analysis because it has no influence on the test of our main hypothesis while increasing model complexity.

should observe earlier fixations on the "Goal-distractor" compared to the "Grip-distractor" if visual attention is first captured by goal information. This would be reflected by a more positive linear estimate or cubic estimate for the goal compared to the grip fixation curve. For example, the cubic time term has been shown to be sensitive to differences in the early and late inflexions of the fixation curves (see Figure 3 of Kalénine et al., 2012 for an illustration). An early increase of fixation proportion on the "Goal-distractor" in comparison to the "Grip-distractor" would thus be statistically reflected by an interaction between the variable "distractor type" and the cubic (3rd polynomial order) time term.

In the main analysis, fixation proportions on the distractor pictures were averaged over items and analysed as a function of the fixed-effect factors of time (fourth-order orthogonal polynomials), distractor type ("Grip-distractor", "Goal-distractor") and the interaction between the two factors. The random structure includes random slopes for participants on each time term[3]. In a complementary analysis, fixation proportions on the distractor pictures were averaged over subjects and analysed as a function of the fixed-effect factors of time, distractor type, their interaction, and image saliency index and its interaction with time. By adding the saliency index covariate to the model, this complementary by-item analysis aimed at partialling out the influence of low-level visual features on the fixation curves. The random structure includes random slopes for items on each time term[4]. Mixed-effect models of fixation proportions were then fitted with REML using the lmer function from the *lme4 1.1-17* package (Bates, Mächler, Bolker, & Walker, 2015) in R version 3.4.4.

Overall main effects and interactions were evaluated with F statistics using the anova function of the *lmerTest 3.0-1* package (Kuznetsova, Brockhoff, & Christensen, 2017). The degrees of

---

[3] The R syntax of the model was: Fixation proportion = (intercept + linear + quadratic + cubic + quartic) * distractor type + (intercept + linear + quadratic + cubic + quartic | participant : distractor type)

[4] The R syntax of the model was: Fixation proportion = (intercept + linear + quadratic + cubic + quartic) * distractor type + (intercept + linear + quadratic + cubic + quartic) * saliency index (intercept + linear + quadratic + cubic + quartic | participant : distractor type)

freedom of the denominator were approximated with the Satterthwaite's method. This method produces acceptable type 1 error rates (Luke, 2017). T-Tests on individual parameter estimates were then used to evaluate the contrasts of interest between distractors.

Results

Main analysis of fixation proportions

Overall, only trials on which the target image was correctly identified were included in the fixation analyses (mean accuracy 91% +/- 28%). As the task was to find the target action, two participants for whom fixations on the target never reached at least 50% of all fixations were considered performing the task correctly but with an atypical visual strategy and were excluded from the analysis. After visual inspection, the time-window of analysis was selected from display onset to 1500ms after display onset, when the averaged target fixation curve reached a first plateau (see Figure 2 and Lee et al., 2013; Mirman et al., 2008 for similar procedure).

The analysis showed no main effect of distractor type ($F_{1,34} = 0.45$, p = .506). This indicates that overall, grip and goal distractors received an equivalent proportion of fixations over the whole 1500 ms time window ("Grip-distractor" mean proportion 0.21; "Goal-distractor" mean proportion 0.22). Importantly, however, a significant interaction was found between distractor type and the cubic (3rd order) time-term ($F_{1,34} = 4.77$, p = .041), reflecting an influence of distractor type on the time course of fixation proportion. The distractor type x cubic time term interaction was driven by an earlier increase of fixation proportion over the "Goal-distractor" in comparison to the "Grip-distractor" (estimate = - 0.13, SE = 0.06), as visible on Figure 2. Distractor type did not interact with any other time terms (all p > .157).

Complementary analysis of fixation proportions with saliency index as covariate

In the complementary by-item analysis including the saliency index, the interaction between distractor type and the cubic (3rd order) time-term was marginally significant ($F_{1,37} = 3.76$, p = .060) after taking the saliency index into account. As previously observed, there was an earlier rise in fixation proportion over the "Goal-distractor" in comparison to the "Grip-distractor" (estimate = - 0.13, SE = 0.06). Importantly, there were no effects involving the saliency index on fixation proportions, neither in isolation (main effect $F_{1,37} = 0.15$, p = .706), nor in interaction with the different time terms (all p > .477). In addition, at the item level, no correlations were found between the amplitude of grip and goal processing early in the time window (extracted from the random cubic estimates for items) and the saliency index ("Grip-distractor" condition: r = 0.37, p = .107; "Goal-distractor" condition: r = -0.2, p = .399). Overall, the complementary analysis indicates that we can be confident that the earlier fixations on goal-distractors cannot be fully explained by the greater visual saliency of the images in this condition.

[INSERT FIGURE 2 ABOUT HERE]

Discussion

The present study aimed at investigating the spontaneous capture of visual attention by grip and goal information. More specifically, we wanted to determine whether visual attention would be preferentially driven towards grip-related or goal-related information. In a visual search task, participants were asked to explore and select the photograph displaying the correct tool use action among action distractors. Gaze movements were used to evaluate to what extent grip-related (same grip as the target, but with a different action goal) and goal-related (same goal as the target but involving a different grip) distractors would capture participants' visual attention before the identification of the target. Visual attention was found preferentially captured by goal-related distractors in comparison to grip-related distractors, but in a time-dependent manner. Visual attention over the goal-related distractors increased in the first part of the visual exploration but decreased in the second part. Thus, observers do not only use goal-related

1   information overall when decoding others' actions, they rely on it first. They disengage their

2   attention from it afterwards to use the other available information.

3   The importance of goals in action processing has been highlighted in several theoretical models

4   (Bach et al., 2014; Cooper et al., 2014; van Elk et al., 2014) and is supported by many

5   experimental arguments (Flanagan et al., 2013; Nicholson et al., 2017; van Elk, Van Schie, &

6   Bekkering, 2008). Predictive approaches go a step further by suggesting that a prediction about

7   others' goals is first needed to make sense of their actual movement kinematics (Kilner, 2011;

8   Kilner et al., 2007). Yet the greater weight of goal information in action decoding is not

9   sufficient to support the "goal first processing" hypothesis, since strong activation of goal

10   information could be derived from the first analysis of visual kinematics (Kilner & Frith, 2008;

11   Tidoni & Candidi, 2016). Data about the time course of processing of grip and goal information

12   are thus particularly needed to directly evaluate predictive approaches of action understanding

13   (Catmur, 2015). The "goal first processing" hypothesis has been indirectly supported by EEG

14   studies showing an early modulation of brain activity as a function of the goals of observed

15   actions (Ortigue, Thompson, Parasuraman, & Grafton, 2009). In a recent behavioural study,

16   we reported more direct evidence in favour of the "goal first" hypothesis. The recognition of

17   visual actions was facilitated after being briefly primed (66 ms) by actions showing the same

18   action goal but not the same action grip (Decroix & Kalénine, 2018). This result demonstrated

19   that goal-related information is used earlier that information about visual kinematics when the

20   task puts minimal requirement on the visuo-attentional system (i.e. central presentation of one

21   action picture at a time). In the present study, we further show that very early in the action

22   recognition process, goal-related information is favoured over visual kinematics when the two

23   dimensions are competing for attention (i.e. visual search of the correct actions). This suggests

24   that the predictive mechanisms at play in action decoding interact with attentional processes in

25   the determination of the temporal dynamics of action processing.

Although the gaze pattern corroborates the "goal first processing" hypothesis, visual attention during the action discrimination task was not solely captured by correct goal-related information but was also influenced by correct kinematic information. The disengagement of visual attention from the goal-related distractor in the second part of the visual exploration provides further evidence for the use of visual kinematics during action recognition. Visual kinematics are indeed known to provide sufficient information to discriminate between two different goals (Cavallo, Koul, Ansuini, Capozzi, & Becchio, 2016), and observers are able to use such information to anticipate the actor's goal (Ansuini et al., 2014; Fischer, Prinz, & Lotz, 2008; Lewkowicz, Quesque, Coello, & Delevoye-Turrell, 2015). Visual kinematics are thus relevant features for understanding the actor's goal. Predictive approaches suggest that visual kinematics are used to test the goal prediction that has been derived from non-motor related information (Donnarumma, Costantini, Ambrosini, Friston, & Pezzulo, 2017; Kilner, 2011; Kilner et al., 2007). Converging evidence suggests that visual kinematics are used to update predictions about the actor's action goal. Motor simulation has been shown to reflect *expected* visual kinematics during the first steps of action observation but *actual* visual kinematics during the last steps of action observation (Cavallo, Bucchioni, Castiello, & Becchio, 2013). Recently, Koul, Soriano, Tversky, Becchio and Cavallo (2019) further showed that actual visual kinematics are used to update the on-going motor simulation as a function of their informativeness. Accordingly, the overall pattern of fixations reported here supports predictive approaches of action recognition, as visual kinematics became more relevant than goal-related information in the second part of the visual exploration.

In our experiment, information about the object is required to perform the task (search for the correct action according to the typical use of the object) and goal correctness is manipulated by changing the orientation of the object (e.g., pen upside down). It is then possible that the early capture of visual attention by goal-related information has been biased by task demands, which

1   orient towards object processing. When looking for the correct action without more instructions,

2   participants may have primarily searched for object information. As the object was present in

3   each condition, the mere presence of the object could not have favoured one type of distractor

4   over another. However, one may wonder whether visual attention might have been primarily

5   drawn towards distractor objects presented in the same correct orientation as the target, which

6   would have favoured distractors sharing the same correct action goal. Yet it is unclear whether

7   the modification of object orientation has changed object familiarity and/or recognition. Indeed,

8   many object exemplars were simply mirror-reversed according to the vertical axis, which makes

9   the presentation of objects in actions containing goal-violations equally visually familiar despite

10   inappropriate for right-handed use (see supplementary materials). In addition, goal priority was

11   confirmed in the by-item analysis that accounts for the possible heterogeneity in the stimulus

12   set. Therefore, we believe that it is relatively unlikely that the present pattern of results can be

13   fully explained by our manipulation of goal information. An interesting direction for future

14   studies would be to dissociate object function (that provides information about the typical goal

15   of the action) and object identity. The same object with alternative functions (e.g., pour or drink

16   from recipient) or two different objects with the same possible function (e.g., meat-knife or

17   box-cutter to cut) could be used for this purpose, although an independent manipulation of the

18   corresponding use gestures might be challenging. Regardless, the role of object identity in

19   deriving goal-related information requires deeper understanding.

20   Although many important theoretical accounts have suggested a key role of object information

21   in deriving predictions about the actors' goal during action decoding (Bach et al., 2014; van Elk

22   et al., 2014), several authors have discussed the scope of such accounts (see for example

23   commentaries of Hommel, 2014; Uithol & Maranesi, 2014). In particular, it remains to

24   determine whether other types of (non-object directed) actions are still processed in a predictive

25   manner. Some results suggest that it is indeed the case (Bach & Schenke, 2017). For example,

Manera, Becchio, Schouten, Bara and Verfaillie (2011) found that the communicative actions of one actor could be used as pieces of information to predict the actions of another actor, even though there was no direct contact between the two actors. If so, goal-priority during action processing may not be the privilege of object-directed actions. Nevertheless, we believe that the role of kinematics versus goal-related information in action decoding can be sensitive to attentional and situational factors. Goal-priority may be nuanced in certain situations. Some results indeed suggest that it may be possible to modify the way actions are spontaneously processed. For example, Pomiechowska and Csibra (2017) found that perception of object-directed actions did not induce mu suppression (i.e., a neurophysiological marker of sensorimotor cortex activity) when actions were preceded by speech, in comparison to the perception of actions in absence of speech. Future studies should then determine whether task demands and situation could bias the spontaneous orientation of visual attention towards kinematics versus goal-related information during the observation of object-directed actions.

Overall, the present study indicates that the visuo-attentional system is first influenced by goal-related information when searching for the correct action among distractors. Although results provide direct support for predictive approaches of action understanding, they might also be incorporated into a broader theoretical framework in which task demands could flexibly bias visual attention towards visual kinematics or non-motor action-related information.

Open Practices Statements: The experiment reported here was not preregistered. Data are available at https://osf.io/ejtay/. The full set of stimuli is available as supplementary materials.

References:

Ambrosini, E., Costantini, M., & Sinigaglia, C. (2011). Grasping with the eyes. *Journal of Neurophysiology*, *106*(3), 1437–1442. https://doi.org/10.1152/jn.00118.2011

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*(7), 1–9. https://doi.org/10.1371/journal.pone.0067916

Ansuini, C., Cavallo, A., Bertone, C., & Becchio, C. (2014). The visible face of intention: Why kinematics matters. *Frontiers in Psychology*, *5*(JUL), 1–6. https://doi.org/10.3389/fpsyg.2014.00815

Bach, P., Nicholson, T., & Hudson, M. (2014). The affordance-matching hypothesis: how objects guide action understanding and prediction. *Frontiers in Human Neuroscience*, *8*(May), 254. https://doi.org/10.3389/fnhum.2014.00254

Bach, P., & Schenke, K. C. (2017). Predictive social perception: Towards a unifying framework from action observation to person knowledge. *Social and Personality Psychology Compass*, *11*(7), e12312. https://doi.org/10.1111/spc3.12312

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants Parse Dynamic Action. *Child Development*, *72*(3), 708–717.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Buresh, J. S., & Woodward, A. L. (2007). Infants track action goals within and across agents. *Cognition*, *104*(2), 287–314. https://doi.org/10.1016/j.cognition.2006.07.001

Catmur, C. (2015). Understanding intentions from actions: Direct perception, inference, and the roles of mirror and mentalizing systems. *Consciousness and Cognition*, 1–8. https://doi.org/10.1016/j.concog.2015.03.012

Cavallo, A., Bucchioni, G., Castiello, U., & Becchio, C. (2013). Goal or movement? Action representation within the primary motor cortex. *European Journal of Neuroscience*, *38*(10), 3507–3512. https://doi.org/10.1111/ejn.12343

Cavallo, A., Koul, A., Ansuini, C., Capozzi, F., & Becchio, C. (2016). Decoding intentions from movement kinematics. *Scientific Reports*, *6*(November), 37036. https://doi.org/10.1038/srep37036

Coello, Y., & Delevoye-Turrell, Y. (2007). Embodiment, spatial categorisation and action. *Consciousness and Cognition*, *16*(3), 667–683. https://doi.org/10.1016/j.concog.2007.07.003

Cooper, R. P., Ruh, N., & Mareschal, D. (2014). The goal circuit model: A hierarchical multi-

route model of the acquisition and control of routine sequential action in humans. *Cognitive Science*, *38*(2), 244–274. https://doi.org/10.1111/cogs.12067

D'Innocenzo, G., Gonzalez, C. C., Nowicky, A. V., Williams, A. M., & Bishop, D. T. (2017). Motor resonance during action observation is gaze-contingent: A TMS study. *Neuropsychologia*, *103*(November 2016), 77–86. https://doi.org/10.1016/j.neuropsychologia.2017.07.017

Decroix, J., & Kalénine, S. (2018). Timing of grip and goal activation during action perception: a priming study. *Experimental Brain Research*, *236*(8), 2411–2426. https://doi.org/10.1007/s00221-018-5309-0

Donaldson, P. H., Gurvich, C., Fielding, J., & Enticott, P. G. (2015). Exploring associations between gaze patterns and putative human mirror neuron system activity. *Frontiers in Human Neuroscience*, *9*(9), 53–58. https://doi.org/10.3389/fnhum.2015.00396

Donnarumma, F., Costantini, M., Ambrosini, E., Friston, K., & Pezzulo, G. (2017). Action perception as hypothesis testing. *Cortex*, *89*, 45–60. https://doi.org/10.1016/j.cortex.2017.01.016

Fischer, M. H., Prinz, J., & Lotz, K. (2008). Grasp cueing shows obligatory attention to action goals. *Quarterly Journal of Experimental Psychology*, *61*(6), 860–868. https://doi.org/10.1080/17470210701623738

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*(6950), 769–771. https://doi.org/10.1038/nature01861

Flanagan, J. R., Rotman, G., Reichelt, A. F., & Johansson, R. S. (2013). The role of observers' gaze behaviour when watching object manipulation tasks: predicting and evaluating the consequences of action. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *368*(1628), 20130063. https://doi.org/10.1098/rstb.2013.0063

Geangu, E., Senna, I., Croci, E., & Turati, C. (2015). The effect of biomechanical properties of motion on infants' perception of goal-directed grasping actions. *Journal of Experimental Child Psychology*, *129*, 55–67. https://doi.org/10.1016/j.jecp.2014.08.005

Hommel, B. (2014). The role of action semantics in action control. Comment on "Action semantics: A unifying conceptual framework for the selective use of multimodal and modality-specific object knowledge" by Michiel van Elk, Hein van Schie and Harold Bekkering. *Physics of Life Reviews*, *11*(2), 259–260. https://doi.org/10.1016/j.plrev.2014.01.010

Hrkać, M., Wurm, M. F., & Schubotz, R. I. (2014). Action observers implicitly expect actors

to act goal-coherently, even if they do not: An fMRI study. *Human Brain Mapping*, *35*(5), 2178–2190. https://doi.org/10.1002/hbm.22319

Hudson, M., Nicholson, T., Ellis, R., & Bach, P. (2016). I see what you say: Prior knowledge of other's goals automatically biases the perception of their actions. *Cognition*, *146*, 245–250. https://doi.org/10.1016/j.cognition.2015.09.021

Hudson, M., Nicholson, T., Simpson, W. A., Ellis, R., & Bach, P. (2016). One step ahead: The perceived kinematics of others' actions are biased toward expected goals. *Journal of Experimental Psychology: General*, *145*(1), 1–7. https://doi.org/10.1037/xge0000126

Humphreys, G. W., Kumar, S., Yoon, E. Y., Wulff, M., Roberts, K. L., & Riddoch, M. J. (2013). Attending to the possibilities of action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1628), 20130059–20130059. https://doi.org/10.1098/rstb.2013.0059

Hunnius, S., & Bekkering, H. (2014). What are you doing? How active and observational experience shape infants' action understanding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*, 20130490. https://doi.org/10.1098/rstb.2013.0490

Jacob, P., & Jeannerod, M. (2005). The motor theory of social cognition: A critique. *Trends in Cognitive Sciences*, *9*(1), 21–25. https://doi.org/10.1016/j.tics.2004.11.003

Kalénine, S., Mirman, D., Middleton, E. L., & Buxbaum, L. J. (2012). Temporal dynamics of activation of thematic and functional knowledge during conceptual processing of manipulable artifacts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1274–1295. https://doi.org/10.1037/a0027626

Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in Cognitive Sciences*, *15*(8), 352–357. https://doi.org/10.1016/j.tics.2011.06.005

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing*, *8*(3), 159–166. https://doi.org/10.1007/s10339-007-0170-2

Kilner, J. M., & Frith, C. D. (2008). Action Observation: Inferring Intentions without Mirror Neurons. *Current Biology*, *18*(1), R32–R33. https://doi.org/10.1016/j.cub.2007.11.008

Koul, A., Soriano, M., Tversky, B., Becchio, C., & Cavallo, A. (2019). The kinematics that you do not expect: Integrating prior information and kinematics to understand intentions. *Cognition*, *182*(October), 213–219. https://doi.org/10.1016/j.cognition.2018.10.006

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lee, C., Middleton, E., Mirman, D., Kalénine, S., & Buxbaum, L. J. (2013). Incidental and context-responsive activation of structure- and function-based action features during object identification. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 257–270. https://doi.org/http://dx.doi.org/10.1037/a0027533

Lee, C., Mirman, D., & Buxbaum, L. J. (2014). Abnormal dynamics of activation of object use information in apraxia: Evidence from eyetracking. *Neuropsychologia*, *59*(1), 13–26. https://doi.org/10.1016/j.neuropsychologia.2014.04.004

Leonetti, A., Puglisi, G., Siugzdaite, R., Ferrari, C., Cerri, G., & Borroni, P. (2015). What you see is what you get: motor resonance in peripheral vision. *Experimental Brain Research*, *233*(10), 3013–3022. https://doi.org/10.1007/s00221-015-4371-0

Lewkowicz, D., Quesque, F., Coello, Y., & Delevoye-Turrell, Y. N. (2015). Individual differences in reading social intentions from motor deviants. *Frontiers in Psychology*, *6*(August), 1–12. https://doi.org/10.3389/fpsyg.2015.01175

Loucks, J., & Pechey, M. (2016). Human Action Perception is Consistent, Flexible, and Orientation Dependent. *Perception*, *45*(11), 1222–1239. https://doi.org/10.1177/0301006616652054

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Manera, V., Becchio, C., Schouten, B., Bara, B. G., & Verfaillie, K. (2011). Communicative interactions improve visual detection of biological motion. *PLoS ONE*, *6*(1). https://doi.org/10.1371/journal.pone.0014594

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. https://doi.org/978-1-4665-8433-4

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494. https://doi.org/10.1016/j.jml.2007.11.006

Mirman, D., & Magnuson, J. S. (2009). Dynamics of activation of semantically similar concepts during spoken word recognition. *Memory and Cognition*, *37*(7), 1026–1039. https://doi.org/10.3758/MC.37.7.1026

Muthukumaraswamy, S. D., & Singh, K. D. (2008). Modulation of the human mirror neuron system during cognitive activity. *Psychophysiology*, *45*(6), 896–905. https://doi.org/10.1111/j.1469-8986.2008.00711.x

Naish, K. R., Reader, A. T., Houston-Price, C., Bremner, A. J., & Holmes, N. P. (2013). To eat or not to eat? Kinematics and muscle activity of reach-to-grasp movements are

influenced by the action goal, but observers do not detect these differences. *Experimental Brain Research*, *225*(2), 261–275. https://doi.org/10.1007/s00221-012-3367-2

Nicholson, T., Roser, M., & Bach, P. (2017). Understanding the goals of everyday instrumental actions is primarily linked to object, not motor-kinematic, information: Evidence from fMRI. *PLoS ONE*, *12*(1), 1–21. https://doi.org/10.1371/journal.pone.0169700

Novack, M. A., Wakefield, E. M., & Goldin-Meadow, S. (2016). What makes a movement a gesture? *Cognition*, *146*, 339–348. https://doi.org/10.1016/j.cognition.2015.10.014

Ocampo, B., & Kritikos, A. (2011). Interpreting actions: The goal behind mirror neuron function. *Brain Research Reviews*, *67*(1–2), 260–267. https://doi.org/10.1016/j.brainresrev.2011.03.001

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Ortigue, S., Thompson, J. C., Parasuraman, R., & Grafton, S. T. (2009). Spatio-temporal dynamics of human intention understanding in temporo-parietal cortex: A combined EEG/fMRI repetition suppression paradigm. *PLoS ONE*, *4*(9). https://doi.org/10.1371/journal.pone.0006962

Perry, A., Troje, N. F., & Bentin, S. (2010). Exploring motor system contributions to the perception of social information: Evidence from EEG activity in the mu/alpha frequency range. *Social Neuroscience*, *5*(3), 272–284. https://doi.org/10.1080/17470910903395767

Pomiechowska, B., & Csibra, G. (2017). Motor activation during action perception depends on action interpretation. *Neuropsychologia*, *105*(October 2016), 84–91. https://doi.org/10.1016/j.neuropsychologia.2017.01.032

Riach, M., Holmes, P. S., Franklin, Z. C., & Wright, D. J. (2018). Observation of an action with a congruent contextual background facilitates corticospinal excitability: A combined TMS and eye-tracking experiment. *Neuropsychologia*, *119*(August), 157–164. https://doi.org/10.1016/j.neuropsychologia.2018.08.002

Schuch, S., Bayliss, A. P., Klein, C., & Tipper, S. P. (2010). Attention modulates motor system activation during action observation: Evidence for inhibitory rebound. *Experimental Brain Research*, *205*(2), 235–249. https://doi.org/10.1007/s00221-010-2358-4

Thill, S., Caligiore, D., Borghi, A. M., Ziemke, T., & Baldassarre, G. (2013). Theories and computational models of affordance and mirror systems: An integrative review. *Neuroscience and Biobehavioral Reviews*, *37*(3), 491–521.

https://doi.org/10.1016/j.neubiorev.2013.01.012

Tidoni, E., & Candidi, M. (2016). Commentary: Understanding intentions from actions: Direct perception, inference, and the roles of mirror and mentalizing systems. *Frontiers in Behavioral Neurosciences*, *10*, 1–3. https://doi.org/10.3389/fnbeh.2016.00013 A

Uithol, S., & Maranesi, M. (2014). No need to match: a comment on Bach, Nicholson and Hudson's "Affordance-Matching Hypothesis." *Frontiers in Human Neuroscience*, *8*(September), 593–609. https://doi.org/10.3389/fnhum.2014.00710

Urgesi, C., Candidi, M., & Avenanti, A. (2014). Neuroanatomical substrates of action perception and understanding: an anatomic likelihood estimation meta-analysis of lesion-symptom mapping studies in brain injured patients. *Frontiers in Human Neuroscience*, *8*(May), 344. https://doi.org/10.3389/fnhum.2014.00344

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review*, *94*(1), 3–15. https://doi.org/10.1037/0033-295X.94.1.3

Vallacher, R. R., & Wegner, D. M. (2012). Action Identification Theory. In *Handbook of Theories of Social Psychology: Volume 1* (pp. 327–348). 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd. https://doi.org/10.4135/9781446249215.n17

van Elk, M., van Schie, H., & Bekkering, H. (2014). Action semantics: A unifying conceptual framework for the selective use of multimodal and modality-specific object knowledge. *Physics of Life Reviews*, *11*(2), 220–250. https://doi.org/10.1016/j.plrev.2013.11.005

van Elk, M., Van Schie, H. T., & Bekkering, H. (2008). Conceptual knowledge for understanding other's actions is organized primarily around action goals. *Experimental Brain Research*, *189*(1), 99–107. https://doi.org/10.1007/s00221-008-1408-7

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, *19*(9), 1395–1407. https://doi.org/10.1016/j.neunet.2006.10.001

Wolpert, D., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *358*(1431), 593–602. https://doi.org/10.1098/rstb.2002.1238

Woodruff, C. C., & Klein, S. (2013). Attentional distraction, μ-suppression and empathic perspective-taking. *Experimental Brain Research*, *229*(4), 507–515. https://doi.org/10.1007/s00221-013-3612-3

Wright, D. J., Wood, G., Franklin, Z. C., Marshall, B., Riach, M., & Holmes, P. S. (2018).

Directing visual attention during action observation modulates corticospinal excitability. *PLoS ONE*, *13*(1), 1–15. https://doi.org/10.1371/journal.pone.0190165

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*(1), 29–58. https://doi.org/10.1037//0096-3445.130.1.29
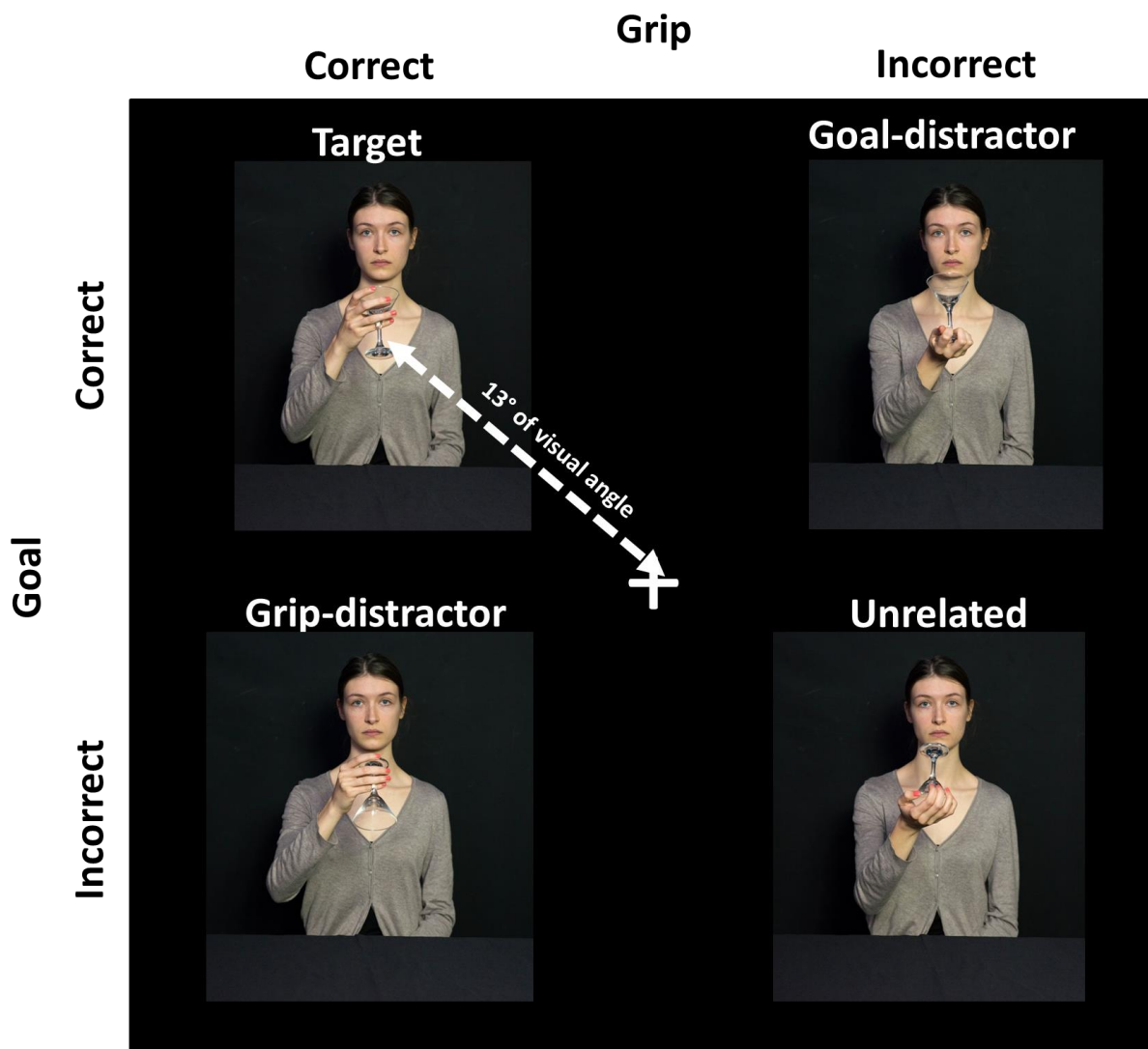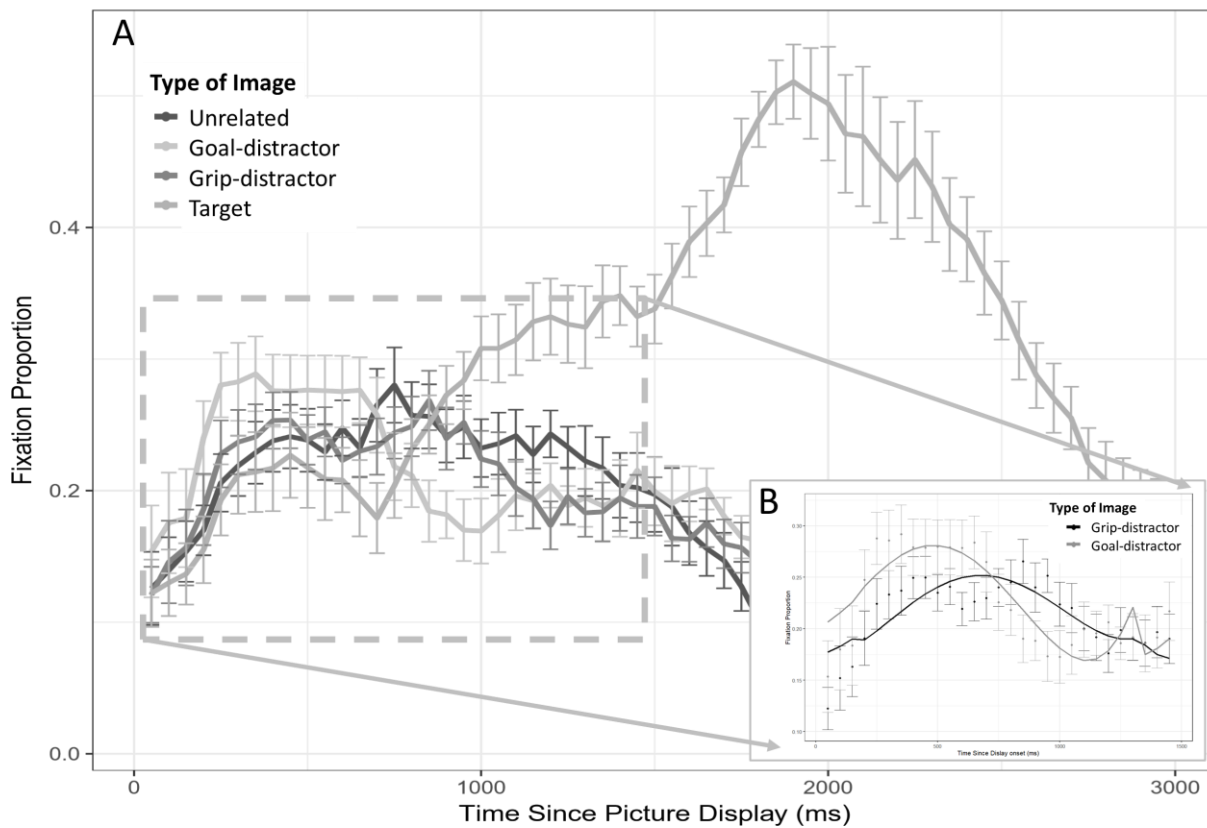
1   **Figure 1.** Design of the experiment.

2   **Figure 2.** Mean fixation proportion and standard errors (error bars) over time as a function of

3   image condition (A) and model fit of the data for the Grip-distractor and Goal-distractor (B)

4

Grip

Correct   **Incorrect**

Target   **Goal-distractor**

Goal — **Correct**

13° of visual angle

Grip-distractor   **Unrelated**

Incorrect