# Body posture differentially impacts on visual attention towards tool, graspable, and non-graspable objects.

Ettore Ambrosini[1] & Marcello Costantini[2,3]

[1] Department of Neuroscience, University of Padua, Via Giustiniani 5, 35128 Padua, Italy

[2] Centre for Brain Science, Department of Psychology, University of Essex, UK

[3] Department of Neuroscience and Institute for Advanced Biomedical Technologies, University "G. d'Annunzio", Chieti, Italy.

**Abbreviated title:** Perception for action

**Corresponding author:** Ettore Ambrosini. Department of Neuroscience, University of Padua, Via Giustiniani 5, 35128 Padua, Italy. E-mail: ettore.ambrosini@gmail.com

**Conflict of interest:** The authors declare no competing financial interests.

**Abstract**

Viewed objects have been shown to afford suitable actions, even in absence of any intention to act. Little is known, however, as to whether gaze behavior, that is the way we simply look at objects, is sensitive to action afforded by the seen object, and how our actual motor possibilities affect this behavior. We recorded participants' eye movements during the observation of tools, graspable and ungraspable objects while their hands were either freely resting on the table or tied behind their back. The effects of the observed object and hand posture on gaze behavior were measured by comparing the actual fixations distribution with that predicted by two widely supported models of visual attention, namely the Graph-Based Visual Saliency and the Adaptive Whitening Salience models. Results showed that saliency models did not predict accurately participants' fixation distributions for tools. Participants, indeed, mostly fixated the action-related, functional part of the tools, regardless of its visual saliency. Critically, the restriction of the participants' action possibility led to a significant reduction of this effect and significantly improved the models prediction of the participants' gaze behavior. We suggest, first, that action-relevant object information at least in part guides gaze behavior. Second, postural information interacts with visual information to the generation of priority maps of fixation behavior. We support the view that the kind of information we access from the environment is constrained by our readiness to act.

## 1. Introduction

On one view, visual perception is a modular encapsulated process that is unaffected by nonvisual factors (Pylyshyn, 2003). On a different view, visual perception is embodied in the sense that it relates body states and goals to the opportunities of acting in the environment (Proffitt, 2006; Proffitt & Linkenauger, 2013). According to the latter view, perception, and particularly object perception, heavily depends upon our possibility to act in the environment (Gibson, 1979). Gibson originally put forward this idea using the notion of affordance. Affordance is defined as the demand to act offered by the environment/objects. But how deeply is action information tied to object perception? And how deeply does our possibility to act impact the way we visually explore objects? To answer these questions we investigated gaze behavior while healthy participants observed common tools (e.g. pliers), non-tools graspable objects (e.g. towel) and ungraspable objects (e.g. barrel).

The correct allocation of visual attention in space and time is mandatory in order to accomplish visually-guided behavior. Indeed, to proficiently interact with the environment, an agent has to attend locations relevant to the ongoing behavioral goal, and this can be done efficiently by directing foveal vision and fixating those locations to extract the relevant information (Land, 2006). Pioneering studies by Koch and Ullman (1985), (see also: Itti & Koch, 2000; Itti, Koch, & Niebur, 1998) have provided reliable models able to predict, from low-level, bottom-up visual features, those locations. Further studies have largely elaborated on these models providing evidence showing that gaze behavior reflects the interplay between bottom-up and top-down sources of information generating priority maps (Kowler, 2011; Malcolm & Henderson, 2010; Tatler et al., 2013; Torralba, Oliva, Castelhano, & Henderson, 2006). This holds true for both complex visual scenes and single objects (Tatler, et al., 2013). Interestingly, action goals, conceived as top-down sources of information, play a pivotal role on the generation of these priority maps (Ballard, Hayhoe, Li, & Whitehead, 1992; Einhauser, Rutishauser, & Koch, 2008a; Rothkopf, Ballard, & Hayhoe, 2007). This is evident in the tight coupling between vision and action during object manipulations, where the selection of priorities depends heavily on the ongoing behavioral goal (Belardinelli, Herbort, & Butz, 2015; Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; Land, Mennie, & Rusted, 1999; Land, 2006). However, regardless of our intentions and goals, graspable objects, especially tools, are intrinsically associated with motor goals and have a specific functional identity (Bub, Masson, & Cree, 2008; Bub & Masson, 2010; Creem-Regehr & Lee, 2005). What is more, even in absence of any intention to act, viewing graspable objects, and in particular tools, triggers suitable motor actions provided that the observer has the actual ability to act (Ambrosini, Sinigaglia, & Costantini, 2012).

Drawing from this knowledge, we investigated gaze behaviour towards everyday tools, non-tools graspable objects, and ungraspable objects. If action information impacts on the way we explore objects, we expect that the pattern of fixations during the observation of tools be mostly focused on object's action-relevant parts whilst the pattern of fixations during the observation of both non-tool graspable and ungraspable objects should not.

Furthermore, we tested the action information effect on visual exploration by manipulating the degree of activation of implicit motor plan elicited by object observation (Ambrosini, Costantini, & Sinigaglia, 2011; Ambrosini, Sinigaglia, et al., 2012; Costantini, Ambrosini, Cardellicchio, & Sinigaglia, 2014; Costantini, Ambrosini, Scorolli, & Borghi, 2011; Costantini, Ambrosini, & Sinigaglia, 2012a, 2012b; Costantini, Ambrosini, Tieri, Sinigaglia, & Committeri, 2010). To this aim, we limited the action ability of a group of participants by tying their hands behind their backs, a manipulation that has proven effective in modulating performance in tasks that recruit motor resources (Ambrosini, Sinigaglia, et al., 2012; Ionta & Blanke, 2009; Ionta, Fourkas, Fiorio, & Aglioti, 2007). Hence, if the supposed bias of the pattern of fixations towards the object's action-relevant parts is due to the recruitment of motor representations pertaining the skillful interaction with them, we expect to observe a shift in the fixations distribution from the action-relevant to the perceptually-salient part of tool pictures when participants were temporarily unable to perform the evoked actions.

## 2. Methods

### 2.1. Participants

Forty healthy undergraduate students took part in the study for course credits. All participants provided informed consent, had normal or corrected-to-normal visual acuity, and were right-handed. The study was carried out in accordance with the Declaration of Helsinki and was approved by the local ethical committee. **The first twenty (15 female, mean age = 21.7 years) participants were assigned to the unconstrained posture condition, the other twenty (13 female, mean age = 21.3 years) were assigned to the constrained posture condition (see Stimuli and Procedure section).**

### 2.2. Apparatus

Participants' eye movements were recorded with a remote infrared eye tracker (RK-826PCI pupil/corneal tracking system; ISCAN ETL-400, Burlington, MA). The eye tracker recorded the position of the right eye during observation of stimulus pictures at a sampling rate of 120 Hz. Stimuli were displayed on a 17-inch LCD monitor (60 Hz refresh rate; 1240 × 1028 pixels screen resolution). The monitor was placed 60 cm in front of the participants and a headrest was used to maintain a constant viewing distance and to prevent head movement.

### 2.3. Stimuli and Procedure

The images used in the experiment consisted of 60 digitized pictures depicting common everyday man-made objects taken from Google Images. The stimuli were rendered in grayscale on a uniform white background, and their scales were standardized within a 500 × 500 pixel frame to subtend about 12.5°. The stimuli were balanced for average pixel brightness and for the number of non-background pixels occupied by each object by using custom scripts written in Matlab (the Mathworks, Inc.). The 60 objects were equally subdivided into three categories: 1) Tools (e.g., pliers), which presents, in a clear distinguishing way, a functional part (e.g., the jaws); 2) Graspable objects (e.g., a sponge), which are small enough to be picked up and usable as a whole; and 3) Ungraspable objects (e.g., a couch), which are too big to be acted upon by hands. Figure 1 shows a sample of the stimulus images.

**Figure 1 Near Here**

Each participant completed two recording blocks, in each of which 30 pictures were presented, balanced for category in a randomized order. Each recording block began with a standard nine point calibration procedure to ensure eye movements were correctly monitored and recorded during the experiment (Ambrosini et al., 2011). Each trial began with a fixation cross, which was presented randomly at either 8° above or below the center of the screen (i.e., outside the area occupied by the objects), and remained visible for 4000 ms. Then, object images were presented centrally for 6000 ms (see Figure 1). Participants were simply asked to observe the images, without any particular constraints apart that to refrain from blinking during the presentation of the object.

During the presentation of the stimuli, half of the participants positioned their hands on the table in front of them in a natural resting position (unconstrained hands condition), while the other half held their hands tied behind their back (constrained hands condition).

### 2.4. **Data analysis**

As a first step, raw gaze traces were pre-processed with an ad-hoc algorithm implemented in Matlab to discard blinks and noisy artifacts and to distinguish saccade jumps (detected using a velocity criteria: point-to-point velocity of the gaze trace > 35 deg/s) from fixations. Therefore, pre-processed fixation gaze data consisted in all those data points that were not categorized as blinks, noise, or saccades. Next, we compared quantitatively the distribution of participants' fixations with that predicted by models of visual saliency. To this aim, we used a slightly modified version of the Fixation Region Overlap Analysis (FROA) methodology (see Johnston & Leek, 2009; Leek et al., 2012; for a full description; see Fig. 2).

In brief, for each object we determined an observed area of interest (oAOI) and two predicted areas of interest (pAOIs). The oAOI was created empirically from participants' pre-processed fixation gaze data. The pAOIs were created using an algorithm, from model-based theoretical predictions, rather than arbitrarily, i.e., on the basis of subjective criteria defined by the researcher (see Caldara & Miellet, 2011, for a discussion of problems arising from the a priori segmentation of the images).

To determine the oAOI, for each object in each experiment, we first applied a 2-D Gaussian smoothing function (*SD* = 0.5°) to the filtered gaze data of each participant. In this way, the oAOI also takes into account the within-and between-subjects variability, as well as measurement errors. Because the number of fixation data points varied between subjects and objects, the resulting smoothed fixation maps were normalized to the 0-1 range (min-max normalization). Next, we created a global fixation map (oMAP) of each object by averaging the normalized fixation maps of the 20 participants in each body posture group and normalizing again the resulting map to 1. Finally, the oAOI was determined, at the group level, by binary thresholding the corresponding oMAP using a fixed parameter (0.5) across all conditions, the oAOIs representing the thresholded region maps for the fixation data (Figure 2). In other words, the oAOI consisted of those areas of the oMAP that exceeded the threshold value of 0.5, and thus showed the highest density of fixation data points. It is important to emphasize that the choice of this threshold does not affect the final result (Johnston & Leek, 2009; Leek, et al., 2012).

After determining the oAOIs, we calculated for each object the pAOIs predicted by two bottom-up visual saliency models, that is, the Graph-Based Visual Saliency (GBVS; Harel, Koch, & Perona, 2006) and the Adaptive Whitening Saliency (AWS, Antón Garcia-Diaz, Fdez-Vidal, Pardo, & Dosil, 2012; Garcia-Diaz, Leboran, Fdez-Vidal, & Pardo, 2012) models. These bottom-up models provide a measure of the saliency of each location in the image, the so-called predicted saliency map (pMAP, see Figure 2), on the basis of various low-level visual features. It should be noted that the GBVS model also takes into account the so-called "image center-bias" (Bindemann, 2010; Tatler, 2007) by promoting higher saliency values in the center of the image plane. Therefore, because objects were presented in the center of the screen, the GBVS model would predict both the image center-bias and potential object center-bias (Henderson, 1993; Nuthmann & Henderson, 2010) or center-of-mass effects (e.g., Vishwanath & Kowler, 2003). Moreover, the AWS has recently been shown to be the best performing model in predicting humans' fixations during the observation of photographs of common natural scenes (Borji, Sihite, & Itti, 2013; see also Stoll, Thrun, Nuthmann, & Einhauser, 2015).

Both saliency maps (GBVS and AWS) were calculated for each object using the Matlab implementation of the corresponding algorithms, and consist of a visual salience value (range: 0-1) for each pixel of the image. This salience value indicates the probability that the corresponding location of the image will be fixated on the basis of its low-level perceptual features. The integral of the pMAP were then approximated to that of the corresponding global fixation map by using the *imhistmatch* function in Matlab, in order to ensure that thresholded areas of interest derived from the saliency models were approximately equivalent in size to those derived from the fixation data. Finally, the pAOIs were determined by binary thresholding the corresponding saliency maps using the same criterion-threshold of 0.5 (Figure 2). These empirical and predicted binary AOI region maps formed the basis for the subsequent analysis of participants' gaze behavior during the observation of our stimuli.

At this point, for each object in each experiment, we evaluated the goodness of the prediction of each of the two saliency models by calculating the "Actual Overlap Percentage" (AOP), defined as the amount of spatial overlap between the oAOI for each stimulus and the pAOI for each saliency model normalized by the size of the oAOI (Figure 2). The statistical significance of the observed overlap percentage is then determined with reference to a critical value, that is the "Chance Overlap Percentage" (COP), which corresponds to the percentage of overlap we would expect at the 95% confidence interval of a random distribution of oAOI-pAOI overlap (Figure 2). The bootstrapped probability distributions were derived from Monte Carlo simulations (1000 iterations). Monte Carlo simulations were ran separately for each stimulus, experiment, and data–

model contrast (Johnston & Leek, 2009; Leek, et al., 2012). Now, by comparing Actual and Chance Overlap Percentage values we were able to determine if the corresponding saliency model reliably predicts the pattern of participants' gaze behavior: AOP values greater than COP values indicate that that model significantly predict fixations distribution.

**Figure 2 Near Here**

To obtain a more sensitive measure of the degree of the correspondence between observed fixation data and predicted saliency maps, we calculated a "Model Matching Dissimilarity" index (MMD) by subtracting AOP from COP values. Therefore, lower (negative) values of MMD indicate better correspondence between the tested model and the observed fixation data (i.e., reliable predictions), whereas higher values of MMD indicate worse observed fixation data-saliency model correspondence. It is important to note that the MMD distance measure is robust against variation in oAOI and pAOI size across items, because both COP and AOP are expressed as percentages of the thresholded fixation map of the corresponding object. The MMD value was the primary dependent variable of our subsequent analyses.

## 3. Results

### 3.1. Models Matching Dissimilarity

We compared MMD values across object categories and body postures to assess the goodness with which the saliency models predicted participants' gaze behavior, and whether the actual state of an observer's body, in terms of her specific action ability (Ambrosini, Sinigaglia, et al., 2012; Mele, 2003), could affect the way we visually explore objects. We ran a mixed-design, by-items ANOVA on MMD values with saliency Model (GBVS vs AWS) and Body Posture (Unconstrained hands vs. Constrained hands) as within-items factors, and Object Category (Tool, Graspable, and Ungraspable objects) as between-items factor.

The ANOVA revealed the marginally significant effects of the Model factor ($F_{1,57} = 3.09$, $p$

= .084, $\eta_p^2$ = .051) and the Object Category by Model interaction ($F_{2, 57}$ = 3.02, $p$ = .057, $\eta_p^2$ = .096): The AWS model tended to predict participants' fixations better than the GBVS model did (5.96% vs. 10.18%, *SD* = 19.37% and 18.65%, respectively), especially for Tool objects (7.58% vs. 18.86%, *SD* = 14.34% and 16.40%, respectively).

Critically, the ANOVA revealed a significant Body Posture by Object Category interaction ($F_{2, 57}$ = 8.24, $p$ < .001, $\eta_p^2$ = .224). This interaction indicates that body posture manipulation was effective in modulating participants' gaze behavior specifically during the observation of Tools. Indeed, the Newman-Keuls's post-hoc tests revealed that when the participants' action possibility was reduced by tying their hands behind their back, the MMD values for Tool objects were lower (9.47%, *SD* = 12.74%) as compared to when the participants' were free to move their hands (16.97%, *SD* = 14.36%; $p$ = .002), indicating a better fixation data-model correspondence (see Figure 3). This effect of the posture modulation was not significant for either the Graspable or the Ungraspable objects (both *p*s > .160).

To further investigate the Body Posture by Object Category interaction, we compared the effect of the Body Posture manipulation on participants' gaze behavior across object categories. We thus computed a difference score by subtracting the mean MMD values in the constrained condition from that in the unconstrained condition and carried out a between-item one-way ANOVA with object category as factor. The Newman-Keuls's post-hoc tests on the Object Category effect revealed that the effect of the Body Posture manipulation was significantly higher for the Tool objects (7.51%, *SD* = 8.65%) as compared to the Graspable (-2.68%, *SD* = 7.43%; $p$ = .001) and Ungraspable (-.69%, *SD* = 9.06%; $p$ = .003) objects. Moreover, the effect of the Body Posture manipulation was reliable for the Tool category only, as revealed by one-sample one-tailed *t*-tests against 0 on the Unconstrained-Constrained MMD difference scores (Tool: $t_{19}$ = 3.88, $p$ < .001, Cohen's $d$ = .868; Graspable: $t_{19}$ = -1.61, $p$ = .062, $d$ = -.360; Ungraspable: $t_{19}$ = -.34, $p$ = .369, $d$ = -.076).

To sum up, these results showed that the restriction of the participants' action possibility led to a significant reduction of the dissimilarity between the model prediction and the participants' gaze behavior specifically during the observation of Tool objects.

**Figure 3 Near Here**

### 3.2. **Spatial and temporal difference of fixations distribution for the tool category**

The analysis of the correspondence between the observed fixation distributions and the models previsions indicated that the way we visually explore tools is influenced by our specific action abilities. Since tools are characterized by spatially separated functional parts (the head of the hammer) and manipulation parts (the handle), we investigated in more details the relative influence of the functional representations that would be activated by the observation of this part on the spatial and temporal distribution of participants' fixations.

To this aim, we first partitioned the entire area occupied by each tool to determine the functional part of the tool and normalized its size by computing the percentage of the total object area occupied by it ($M$ = 54.1%, $SD$ = 21.0%). Next, for each participant and object, we calculated the percentage of the entire set of pre-processed, filtered data points (excluding those that were not located within the area occupied by the object) that were located within the functional part. This procedure was performed 1) for each 500 ms bin of the entire presentation time (6000 ms), and 2) for the first five fixations. We then normalized this percentage values by subtracting the percentage of the area occupied by the functional part from it, obtaining a normalized percentage (norm%) of the fixation gaze data located within the functional part of the tool. From now on, we refer to this measure as Normalized Fixation Functional (NFF). Therefore, the resulting NFF values take into account variation in the size of the functional part across tools, and represents the degree with which the observed fixations distribution exceed the distribution that one would expect by chance. In the same way, we also calculated the percentage of fixation data points that were located within the visually salient part of the tool, that is, the pAOIs predicted by the GBVS and the AWS models (see Section 2.4 and Figure 2) in each 500 ms bin and for each of the first five fixations. Again, for both the GBVS and AWS saliency models, we normalized these percentage values for the size of the corresponding pAOI as described above. We thus obtained the norm% values for the visually salient part of the tools (hereafter Normalized Fixation Saliency, NFS) as predicted by the GBVS and AWS models (respectively, $NFS_{GBVS}$ and $NFS_{AWS}$); these measures can be safely compared to the NFF one.

Finally, the difference in the spatial distributions of fixations occurring within the functional and salient part of each tool, as well the strength of the action possibility modulation of these distributions, were assessed over time and fixations. We did this by carrying out two mixed-design, by-subjects repeated-measure ANOVA on the norm% values with Body Posture (Unconstrained vs. Constrained) as between-subjects factor and Tool Part (Functional vs. GBVS-Salient vs. AWS-Salient) and either Time bin (12 levels, from 500 to 6000 ms) or Fixation (5 levels, from the 1$^{st}$ to the 5$^{th}$ fixation) as within-subjects factors. Post-hoc Newman-Keuls test was used when necessary. When the sphericity assumption was violated, Huynh-Feldt corrected degrees of freedom were reported for the *F* statistic.

### ***3.2.1. Time bins***

The ANOVA revealed a significant main effect of the Time bin factor ($F_{6.37, 242.19} = 3.45$, $p = .002$, $\eta_p^2 = .157$) and a marginally significant effect of the Tool Part factor ($F_{2, 76} = 2.64$, $p = .078$, $\eta_p^2 = .065$), which were further qualified by their significant interaction ($F_{12.07, 458.70} = 9.70$, $p < .0001$, $\eta_p^2 = .203$) (see Figure 4A). Post-hoc analysis showed that, NFF values were higher during the first 1000 ms compared to all the other time bins (19.08% and 20.79% for 500 and 1000 ms bins, respectively; all *p*s < .017) and during the 1500 ms time bin (14.74%) as compared to all but the 2000 and 6000 ms time bins (all *p*s < .026). In addition, the $NFS_{GBVS}$ value for the first time bin was higher than those for the 1000 and 1500 ms bins (10.35% vs. 2.17% and 2.46%, respectively; all *p*s = .002) and the $NFS_{AWS}$ value for the first time bin was lower than those for the 1000, 1500, and 2000 ms bins (3.73% vs. 12.41%, 12.46%, and 12.24%, respectively; all *p*s < .002). Critically, NFF values were significantly higher than the $NFS_{GBVS}$ values during the first 1500 ms of object presentation (all *p*s < .001), and they were also higher than the $NFS_{AWS}$ values in the first 1000 ms (all *p*s < .001) (see Figure 4A).

The ANOVA also revealed the significant Tool Part by Body Posture interaction ($F_{2, 76}$ = 5.19, $p$ = .008, $\eta_p^2$ = .120, see Figure 4B). Post-hoc analysis revealed that, on average, NFF values were significantly higher than $NFS_{GBVS}$ values (i.e., the Functional parts of the tools were more fixated than GBVS-Salient ones) when the participants' were free to move their hands (13.85% vs. 4.70%, respectively; $p$ = .004) and this difference was significantly higher than the non-significant one found in the Constrained condition (6.34% vs. 8.13%, respectively; $p$ = .459). Moreover, the NFF values were significantly higher in the Unconstrained as compared to the Constrained condition ($p$ = .039). No other effects were significant[1].

#### *3.2.2. __Fixations__*

The ANOVA revealed a significant main effect of the Fixation and Tool Part factors ($F_{4, 152}$ = 4.53, $p$ = .002, $\eta_p^2$ = .106; $F_{1.75, 66.31}$ = 2.64, $p$ = .025, $\eta_p^2$ = .098, respectively), which were further qualified by their significant interaction ($F_{6.64, 252.49}$ = 6.29, $p$ < .0001, $\eta_p^2$ = .142) (see Figure 4C). Post-hoc analysis showed that, NFF values were higher for the first two fixations (14.43% and 12.11%, respectively) as compared to the 4$^{th}$ and 5$^{th}$ ones (6.92% and 3.89%, respectively; all *p*s < .019) and for the 3$^{rd}$ fixation (12.11%) as compared to the 5$^{th}$ one ($p$ = .008). No differences were found for the $NFS_{GBVS}$ values across fixations, while the $NFS_{AWS}$ value for the 2$^{nd}$ fixation was higher than those for the 1$^{st}$ and 5$^{th}$ ones (13.16% vs. .74% and 3.47%, respectively; all *p*s < .010), and the $NFS_{AWS}$ value for the 3$^{rd}$ fixation (9.27%) was higher than that for the 1$^{st}$ one ($p$ =.098). Critically, NFF values were significantly higher than the $NFS_{GBVS}$ values for the 2$^{nd}$ and 3$^{rd}$ fixations (all *p*s < .009), and they were also higher than the $NFS_{AWS}$ values for the 1$^{st}$ fixation ($p$ <

[1] We also carried out a similar ANOVA by excluding the pAOI predicted by the GBVS model (i.e., the GBVS-Salient level of the Tool Part factor), as the previous analysis of the correspondence between the observed fixation distributions and the models previsions indicated that this model tended to predict less accurately the participants' gaze behavior as compared to the AWS model, especially for the tools. The reported results were essentially the same.

.001) (see Figure 4C).

The ANOVA also revealed the significant Tool Part by Body Posture interaction ($F_{1.75, 66.31}$ = 4.71, $p$ = .025, $\eta_p^2$ = .110, see Figure 4D). Post-hoc analysis revealed that, on average, NFF values were significantly higher than both $NFS_{GBVS}$ and $NFS_{AWS}$ values (i.e., the Functional parts of the tools were more fixated than the visually salient ones) when the participants' were free to move their hands (14.48% vs. 4.78% and 7.62%, respectively; all $p$s < .006) and these differences were significantly higher (respectively, $p$ = .017 and .037) than the non-significant ones found in the Constrained condition (6.38% vs. 6.83% and 6.19%, respectively; $p$ = .852 and .934). Moreover, the NFF values were significantly higher in the Unconstrained as compared to the Constrained condition ($p$ = .007). No other effects were significant[2].

Taken together, the results of the analyses of the spatio-temporal differences of fixations distributions for the tool category confirm and refine those of the previous analyses, showing that the participants' gaze behavior during the observation of Tool objects, especially for the first fixation or time bins, was mostly focused on their functional part and, thus, was not accurately predicted by saliency models. Moreover, they confirm that the way we look at tools depends on our specific action abilities.

**Figure 4 Near Here**

## 4. Discussion

We investigated whether gaze behavior towards everyday tools is sensitive to the goal we can accomplish with them and how our actual motor possibilities affect this behavior. We recorded participants' eye movements during the observation of tools, graspable, and ungraspable objects while their hands were either freely resting on the table (Unconstrained hands) or tied behind their back (Constrained hands). The effects of the observed object (Tool vs. Graspable vs. Ungraspable)

---

[2] Again, we also carried out a similar ANOVA by excluding the pAOI predicted by the GBVS model (i.e., the GBVS-Salient level of the Tool Part factor). The results were essentially the same.

and hand posture (Unconstrained Vs. Constrained) on gaze behavior were measured by comparing the actual fixations distribution with that predicted by two accredited models of visual exploration, namely the Graph-Based Visual Saliency (GBVS) model (Harel, et al., 2006) and the Adaptive Whitening Saliency (AWS) model (Garcia-Diaz et al., 2012a, b).

Both models did not predict accurately fixation distributions for tools[3]. Participants, indeed, fixated the functional part of the tools (Bub, et al., 2008) regardless of the visual saliency, especially for the first fixation or time bins. This suggests that the functional knowledge of the stimulus affected gaze behavior towards tools (Roberts & Humphreys, 2011). This effect was significantly reduced when participants had their hands tied behind their backs. We suggest that the actual possibility to act upon an object, which is not taken into account by visual saliency models, at least in part guide gaze behavior. How can we account for these findings?

One possibility is to look at those studies showing an effect of action knowledge or intention on object representation and recognition. For example, it has been shown that a specific action intention can bias visual processing of action-related objects and visual features (Bekkering & Neggers, 2002; Gutteling, Kenemans, & Neggers, 2011; Symes, Tucker, Ellis, Vainio, & Ottoboni, 2008). Moreover, neuropsychological evidence showed that action templates activated by functional affordances may influence visual search and selection independently of their perceptual properties (Humphreys & Riddoch, 2001). Here we took advantage from the fact that representation of tools is grounded within the sensory-motor system, and tools observation recruits action representations (Matheson, White, & McMullen, 2015). This is supported by numerous behavioral and neural studies showing that observation of objects, particularly tools, induces the covert execution of motor actions (e.g. Tucker & Ellis, 2004; for a review see Martin, 2007). **On the behavioral side, studies on compatibility effects showed that observing pictures of objects or real objects potentiates specific motoric representation of actions, that is the reaching and grasping**

**[3] The data also replicated a pilot study in which only two object categories, i.e. non-graspable object and tool objects, were used.**

**actions we typically perform to pick up and use them for their intended purpose** (Bub, et al., 2008; Tucker & Ellis, 1998, 2001), but only when they afford the potential to be readily used for functional actions (Ambrosini & Costantini, 2013; Ambrosini, Scorolli, Borghi, & Costantini, 2012; Cardellicchio, Sinigaglia, & Costantini, 2011; Costantini, et al., 2014; Costantini, Ambrosini, Scorolli, et al., 2011; Costantini, et al., 2012a, 2012b; Costantini, Ambrosini, Sinigaglia, & Gallese, 2011; Costantini, et al., 2010; Costantini & Sinigaglia, 2012; Ferri, Riggio, Gallese, & Costantini, 2011; Masson, Bub, & Breuer, 2011). These results reveal that manipulable objects are represented in terms of actions that can be realistically executed with them.

Supporting these behavioral and neuropsychological findings, neurophysiological evidence showed that the simple observation of graspable objects leads to the activation of the canonical neuron system (Bonini, Maranesi, Livi, Fogassi, & Rizzolatti, 2014; Murata et al., 1997). The category of artifacts, and particularly tools, can be somewhat peculiar. Indeed, compared to ungraspable objects, observation of tools activates a specific, left-lateralized neural network regardless of the observer's action intention. Along with posterior temporal areas involved in the processing of visual motion (Beauchamp & Martin, 2007), this network includes motor-related brain areas, especially the left premotor and posterior parietal cortices (e.g., Chao & Martin, 2000; Creem-Regehr & Lee, 2005). The activation of this dorsal network when viewing tools would reflect the activation of motor routines for the possible interactions with tools and is considered the neural substrate of affordances (Grezes & Decety, 2002; Jeannerod, 1995).

Thus, behavioral, neurophysiological, and brain imaging studies have demonstrated that seeing objects activates motor representations of their skillful use. Here we propose that such motor recruitment impacts also on the way we simply look at objects. But why did tying participants' hands behind their backs reduce this effect? **One possible explanation pertains the idea that effective observation of tool depends on how readily the motor representation of that tool can be recruited (Ambrosini & Costantini, 2013; Ambrosini, Scorolli, et al., 2012; Cardellicchio, et al., 2011; Costantini, et al., 2014; Costantini, Ambrosini, Scorolli, et al., 2011; Costantini, et**

**al., 2012a, 2012b; Costantini, Ambrosini, Sinigaglia, et al., 2011; Costantini, et al., 2010; Costantini & Sinigaglia, 2012; Ferri, et al., 2011; Masson, et al., 2011). This idea is in line with previous evidence showing that observers' motor abilities are needed for processing others' actions. They show that the richer one's motor repertoire, the greater one's ability to make sense of others' behavior (Aglioti, Cesari, Romani, & Urgesi, 2008; Ambrosini et al., 2013; Calvo-Merino, Glaser, Grezes, Passingham, & Haggard, 2005; Cross, Hamilton, & Grafton, 2006). These findings could be explained by the action-specific perception account (Witt, 2011), according to which people perceive the surrounding environment in terms of their ability.** At the neural level, the integration of visual and proprioceptive/postural information might occur in the posterior parietal cortex and/or the superior colliculus, which receives input from a number of non-visual systems (Abrahams & Rose, 1975).

One may possibly argue that the effect we found could also be explained as a body-parts position effect, rather than an action-possibility effect. Indeed, it has been shown that variations in hand position might affect visual processing (Abrams & Weidler, 2014; Brockmole, Davoli, Abrams, & Witt, 2013; Davoli & Brockmole, 2012; Davoli, Brockmole, & Goujon, 2012; Kelly & Brockmole, 2014; Reed, Grubb, & Steele, 2006) and gaze behavior (Thura, Hadj-Bouziane, Meunier, & Boussaoud, 2008). However, this explanation cannot fully account for the fact that our experimental manipulation specifically affected participants' gaze behavior towards tools. Moreover, it has been shown that the hand position effect on object perception is actually action-dependent (Chan, Peterson, Barense, & Pratt, 2013).

Our results complement and extend previous studies on fixation behavior showing that visual exploration involves both low- and high-level information in scenes (van der Linden, Mathot, & Vitu, 2015). A common finding is that what we expect the target to look like and where we expect to find it are important sources of information in gaze behavior (Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009; Kanan, Tong, Zhang, & Cottrell, 2009; Spotorno, Malcolm, & Tatler, 2014; Tatler, Hayhoe, Land, & Ballard, 2011; Torralba, et al., 2006). Interestingly, in our case, the high-level information was intrinsic to the observed objects, which are known to be represented in terms of the action they afford.

According to the visual salience hypothesis, gaze control is a reaction to the visual properties of the stimulus confronting the viewer: we look at scene locations on the basis of image properties, such as intensity, color, and edge orientation, generated in a bottom-up manner from the scene (Harel, et al., 2006; Itti & Koch, 2000; Itti, et al., 1998; Kanan, et al., 2009; Koch & Ullman, 1985; Parkhurst, Law, & Niebur, 2002; Tatler, Baddeley, & Gilchrist, 2005). This hypothesis has had a large impact on research in scene perception, in part because it has been instantiated within a neurobiologically plausible computational model (Itti & Koch, 2000) that has been found to capture gaze behavior under some conditions (e.g. Derrick Parkhurst, Klinton Law, & Ernst Niebur, 2002). Recently, the model proposed by Itti and Koch has been extended to take into account other low-level factors, such as the so-called object center-bias, the tendency to look at the center of objects when observing visual scenes (Henderson, 1993; Nuthmann & Henderson, 2010) and at the center-of-mass of an isolated visual object (e.g., Vishwanath & Kowler, 2003), or the so-called image center-bias, the tendency to look towards the center of images

Despite the prominence of feature-based accounts of eye guidance in recent years, empirical evaluations of such models have shown that these are insufficient to account for human fixation behavior ( Henderson, Brockmole, Castelhano, & Mack, 2007; e.g., Tatler, et al., 2005; 2006). Even the above mentioned extensions of earlier models, such as the GBVS and AWS we used, still showed large gap compared to the human performance, especially when the behavioral task is manipulated, (Einhauser, Rutishauser, & Koch, 2008b; Foulsham & Underwood, 2008; Underwood & Foulsham, 2006; Geoffrey Underwood, Foulsham, van Loon, Humphreys, & Bloyce, 2006). Even if the low, but significant, explanatory power of visual saliency models may account for our results, our interest was not in their explanatory power per se, rather how the observed object (Tool vs. Graspable vs. Ungraspable) and body posture (Unconstrained vs. Constrained) impacted on the way we explore visual objects.

To conclude, the present findings suggest that the way we visually explore object is biased towards action-relevant information (Handy, Grafton, Shroff, Ketay, & Gazzaniga, 2003; Roberts & Humphreys, 2011), and the kind of information we access from them is constrained by our readiness to act.

## References

Abrahams, V. C., & Rose, P. K. (1975). The spinal course and distribution of fore and hind limb muscle afferent projections to the superior colliculus of the cat. *J Physiol, 247*(1), 117-130.

Abrams, R. A., & Weidler, B. J. (2014). Trade-offs in visual processing for stimuli near the hands. *Atten Percept Psychophys, 76*(2), 383-390. doi: 10.3758/s13414-013-0583-1

Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nat Neurosci, 11*(9), 1109-1116.

Ambrosini, E., & Costantini, M. (2013). Handles lost in non-reachable space. *Exp Brain Res, 229*(2), 197-202. doi: 10.1007/s00221-013-3607-0

Ambrosini, E., Costantini, M., & Sinigaglia, C. (2011). Grasping with the eyes. *Journal of Neurophysiology, 106*(3), 1437-1442. doi: 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), e67916. doi: 10.1371/journal.pone.0067916

Ambrosini, E., Scorolli, C., Borghi, A. M., & Costantini, M. (2012). Which body for embodied cognition? Affordance and language within actual and perceived reaching space. *Consciousness and Cognition, Jul 23. [Epub ahead of print]*. doi: 10.1016/j.concog.2012.06.010

Ambrosini, E., Sinigaglia, C., & Costantini, M. (2012). Tie my hands, tie my eyes. *Journal of Experimental Psychology: Human Perception and Performance, 38*(2), 263-266. doi: 10.1037/a0026570

Ballard, D. H., Hayhoe, M. M., Li, F., & Whitehead, S. D. (1992). Hand-eye coordination during sequential tasks. *Philos Trans R Soc Lond B Biol Sci, 337*(1281), 331-338; discussion 338-339. doi: 10.1098/rstb.1992.0111

Beauchamp, M. S., & Martin, A. (2007). Grounding object concepts in perception and action:

evidence from fMRI studies of tools. *Cortex, 43*(3), 461-468.

Bekkering, H., & Neggers, S. F. (2002). Visual search is modulated by action intentions. *Psychol Sci, 13*(4), 370-374.

Belardinelli, A., Herbort, O., & Butz, M. V. (2015). Goal-oriented gaze strategies afforded by object interaction. *Vision Res, 106*, 47-57. doi: 10.1016/j.visres.2014.11.003

Bindemann, M. (2010). Scene and screen center bias early eye movements in scene viewing. *Vision Res, 50*(23), 2577-2587. doi: 10.1016/j.visres.2010.08.016

S0042-6989(10)00402-5

Bonini, L., Maranesi, M., Livi, A., Fogassi, L., & Rizzolatti, G. (2014). Space-dependent representation of objects and other's action in monkey ventral premotor grasping neurons. *J Neurosci, 34*(11), 4108-4119. doi: 10.1523/JNEUROSCI.4187-13.2014

Borji, A., Sihite, D. N., & Itti, L. (2013). Objects do not predict fixations better than early saliency: a re-analysis of Einhauser et al.'s data. *J Vis, 13*(10), 18. doi: 10.1167/13.10.18

Brockmole, J. R., Davoli, C. C., Abrams, R. A., & Witt, J. K. (2013). The World Within Reach: Effects of Hand Posture and Tool Use on Visual Cognition. *Current Directions in Psychological Science, 22*(1), 38-44. doi: 10.1177/0963721412465065

Bub, D. N., Masson, M. E., & Cree, G. S. (2008). Evocation of functional and volumetric gestural knowledge by objects and words. *Cognition, 106*(1), 27-58.

Bub, D. N., & Masson, M. E. J. (2010). Grasping beer mugs: On the dynamics of alignment effects induced by handled objects. *Journal of Experimental Psychology: Human Perception and Performance, 36*(2), 341.

Caldara, R., & Miellet, S. (2011). iMap: a novel method for statistical fixation mapping of eye movement data. *Behav Res Methods, 43*(3), 864-878. doi: 10.3758/s13428-011-0092-x

Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: an FMRI study with expert dancers. *Cereb Cortex, 15*(8), 1243-1249.

Cardellicchio, P., Sinigaglia, C., & Costantini, M. (2011). The space of affordances: A TMS study. *Neuropsychologia, 49*(5), 1369.

Chan, D., Peterson, M. A., Barense, M. D., & Pratt, J. (2013). How action influences object perception. *Front Psychol, 4*, 462. doi: 10.3389/fpsyg.2013.00462

Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage, 12*(4), 478-484. doi: 10.1006/nimg.2000.0635

Costantini, M., Ambrosini, E., Cardellicchio, P., & Sinigaglia, C. (2014). How your hand drives my eyes. *Social Cognitive and Affective Neuroscience, 9*(5), 705-711.

Costantini, M., Ambrosini, E., Scorolli, C., & Borghi, A. (2011). When objects are close to me: Affordances in the peripersonal space. *Psychonomic Bulletin & Review, 18*(2), 302-308.

Costantini, M., Ambrosini, E., & Sinigaglia, C. (2012a). Does how I look at what you're doing depend on what I'm doing? *Acta Psychologica, 141*(2), 199-204. doi: 10.1016/j.actpsy.2012.07.012

Costantini, M., Ambrosini, E., & Sinigaglia, C. (2012b). Out of your hand's reach, out of my eyes' reach. *The Quarterly Journal of Experimental Psychology, 65*(5), 848-855. doi: 10.1080/17470218.2012.679945

Costantini, M., Ambrosini, E., Sinigaglia, C., & Gallese, V. (2011). Tool-use observation makes far objects ready-to-hand. *Neuropsychologia, 49*(9), 2658-2663. doi: 10.1016/j.neuropsychologia.2011.05.013

Costantini, M., Ambrosini, E., Tieri, G., Sinigaglia, C., & Committeri, G. (2010). Where does an object trigger an action? An investigation about affordances in space. *Experimental Brain Research, 207*(1), 95.

Costantini, M., & Sinigaglia, C. (2012). *Grasping affordance: a window onto social cognition. In Joint Attention: New Developments* (Axel Seemann ed.). Cambridge MA: MIT press.

Creem-Regehr, S. H., & Lee, J. N. (2005). Neural representations of graspable objects: are tools special? *Brain Res Cogn Brain Res, 22*(3), 457-469.

Cross, E. S., Hamilton, A. F., & Grafton, S. T. (2006). Building a motor simulation de novo: observation of dance by dancers. *Neuroimage, 31*(3), 1257-1267.

Davoli, C. C., & Brockmole, J. R. (2012). The hands shield attention from visual interference. *Atten Percept Psychophys, 74*(7), 1386-1390. doi: 10.3758/s13414-012-0351-7

Davoli, C. C., Brockmole, J. R., & Goujon, A. (2012). A bias to detail: how hand position modulates visual learning and visual memory. *Mem Cognit, 40*(3), 352-359. doi: 10.3758/s13421-011-0147-3

Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modeling Search for People in 900 Scenes: A combined source model of eye guidance. *Vis cogn, 17*(6-7), 945-978. doi: 10.1080/13506280902834720

Einhauser, W., Rutishauser, U., & Koch, C. (2008a). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *J Vis, 8*(2), 2 1-19. doi: 10.1167/8.2.2

Einhauser, W., Rutishauser, U., & Koch, C. (2008b). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli *J Vis* (Vol. 8, pp. 2 1-19). United States.

Ferri, F., Riggio, L., Gallese, V., & Costantini, M. (2011). Objects and their nouns in peripersonal space. *Neuropsychologia, 49*(13), 3519-3524. doi: 10.1016/j.neuropsychologia.2011.09.001

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *J Vis, 8*(2), 6 1-17. doi: 10.1167/8.2.6

Garcia-Diaz, A., Fdez-Vidal, X. R., Pardo, X. M., & Dosil, R. (2012). Saliency from hierarchical adaptation through decorrelation and variance normalization. *Image and Vision Computing, 30*(1), 51-64. doi: 10.1016/j.imavis.2011.11.007

Garcia-Diaz, A., Leboran, V., Fdez-Vidal, X. R., & Pardo, X. M. (2012). On the relationship between optical variability, visual saliency, and eye fixations: a computational approach. *J*

*Vis, 12*(6), 17. doi: 10.1167/12.6.17

Gibson, J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton-Mifflin.

Grezes, J., & Decety, J. (2002). Does visual perception of object afford action? Evidence from a neuroimaging study. *Neuropsychologia, 40*(2), 212-222.

Gutteling, T. P., Kenemans, J. L., & Neggers, S. F. (2011). Grasping preparation enhances orientation change detection. *PLoS ONE, 6*(3), e17675. doi: 10.1371/journal.pone.0017675

Handy, T. C., Grafton, S. T., Shroff, N. M., Ketay, S., & Gazzaniga, M. S. (2003). Graspable objects grab attention when the potential for action is recognized. *Nat Neurosci, 6*(4), 421-427.

Harel, J., Koch, C., & Perona, P. (2006). *Graph-based visual saliency.* Paper presented at the Advances in neural information processing systems.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *J Vis, 3*(1), 49-63. doi: 10.1167/3.1.6

Henderson, J. M. (1993). Eye movement control during visual object processing: effects of initial fixation position and semantic constraint. *Can J Exp Psychol, 47*(1), 79-98.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. *Eye movements: A window on mind and brain*, 537-562.

Humphreys, G. W., & Riddoch, M. J. (2001). Detection by action: neuropsychological evidence for action-defined templates in search. *Nat Neurosci, 4*(1), 84-88. doi: 10.1038/82940

Ionta, S., & Blanke, O. (2009). Differential influence of hands posture on mental rotation of hands and feet in left and right handers. *Experimental Brain Research, 195*(2), 207.

Ionta, S., Fourkas, A. D., Fiorio, M., & Aglioti, S. M. (2007). The influence of hands posture on mental rotation of hands and feet. *Exp Brain Res, 183*(1), 1-7.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10–12), 1489-1506. doi: 10.1016/s0042-6989(99)00163-7

Itti, L., Koch, C., & Niebur, E. (1998). A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Trans. Pattern Anal. Mach. Intell., 20*(11), 1254-1259. doi: 10.1109/34.730558

Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia, 33*, 1419-1432.

Johnston, S., & Leek, C. (2009). Fixation Region Overlap: A quantitative method for the analysis of fixational eye movement patterns. *Journal of Eye Movement Research, 1(3)*, 1-12.

Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Vis cogn, 17*(6-7), 979-1003. doi: 10.1080/13506280902771138

Kelly, S. P., & Brockmole, J. R. (2014). Hand proximity differentially affects visual working memory for color and orientation in a binding task. *Front Psychol, 5*, 318. doi: 10.3389/fpsyg.2014.00318

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Hum Neurobiol, 4*(4), 219-227.

Kowler, E. (2011). Eye movements: the past 25 years. *Vision Res, 51*(13), 1457-1483. doi: 10.1016/j.visres.2010.12.014

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28*(11), 1311-1328.

Land, M. F. (2006). Eye movements and the control of actions in everyday life *Prog Retin Eye Res* (Vol. 25, pp. 296-324). England.

Leek, E. C., Cristino, F., Conlan, L. I., Patterson, C., Rodriguez, E., & Johnston, S. J. (2012). Eye movement patterns during the recognition of three-dimensional objects: preferential fixation of concave surface curvature minima. *J Vis, 12*(1), 7. doi: 10.1167/12.1.7

Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *J Vis, 10*(2), 4 1-11. doi: 10.1167/10.2.4

Martin, A. (2007). The representation of object concepts in the brain. *Annu Rev Psychol, 58*, 25-45.

Masson, M. E., Bub, D. N., & Breuer, A. T. (2011). Priming of reach and grasp actions by handled objects. *J Exp Psychol Hum Percept Perform, 37*(5), 1470-1484. doi: 10.1037/a0023509

Matheson, H., White, N., & McMullen, P. (2015). Accessing embodied object representations from vision: A review. *Psychol Bull, 141*(3), 511-524. doi: 10.1037/bul0000001

Mele, A. R. (2003). Agents' Abilities. *Noûs, 37*(3), 447.

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V., & Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area F5) of the monkey. *J Neurophysiol, 78*(4), 2226-2230.

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *J Vis, 10*(8), 20. doi: 10.1167/10.8.20

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42*(1), 107-123. doi: 10.1016/s0042-6989(01)00250-4

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention *Vision Res* (Vol. 42, pp. 107-123). England.

Proffitt, D. R. (2006). Embodied perception and the economy of action. *Perspectives on Psychological Science, 1*(2), 110-122.

Proffitt, D. R., & Linkenauger, S. A. (2013). *Perception viewed as a phenotypic expression* (In W. Prinz, M. Beisert, & A. Herwig (Eds.), Tutorials In Action Science ed.): MIT Press.

Pylyshyn, Z. (2003). Return of the mental image: are there really pictures in the brain? *Trends Cogn Sci, 7*(3), 113-118.

Reed, C. L., Grubb, J. D., & Steele, C. (2006). Hands up: attentional prioritization of space near the hand. *J Exp Psychol Hum Percept Perform, 32*(1), 166-177. doi: 10.1037/0096-1523.32.1.166

Roberts, K. L., & Humphreys, G. W. (2011). Action-related objects influence the distribution of visuospatial attention. *Q J Exp Psychol (Hove), 64*(4), 669-688. doi: 10.1080/17470218.2010.520086

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *J Vis, 7*(14), 16 11-20. doi: 10.1167/7.14.16

Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2014). How context information and target information guide the eyes from the first epoch of search in real-world scenes. *J Vis, 14*(2). doi: 10.1167/14.2.7

Stoll, J., Thrun, M., Nuthmann, A., & Einhauser, W. (2015). Overt attention in natural scenes: objects dominate features. *Vision Res, 107*, 36-48. doi: 10.1016/j.visres.2014.11.006

Symes, E., Tucker, M., Ellis, R., Vainio, L., & Ottoboni, G. (2008). Grasp preparation improves change detection for congruent objects. *J Exp Psychol Hum Percept Perform, 34*(4), 854-871. doi: 10.1037/0096-1523.34.4.854

Tatler, B. W. (2007). The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions. *J Vis, 7*(14), 4 1-17. doi: 10.1167/7.14.4

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time *Vision Res* (Vol. 45, pp. 643-659). England.

Tatler, B. W., Baddeley, R. J., & Vincent, B. T. (2006). The long and the short of it: spatial statistics at fixation vary with saccade amplitude and task *Vision Res* (Vol. 46, pp. 1857-1862). England.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: reinterpreting salience. *J Vis, 11*(5), 5. doi: 10.1167/11.5.5

Tatler, B. W., Hirose, Y., Finnegan, S. K., Pievilainen, R., Kirtley, C., & Kennedy, A. (2013). Priorities for selection and representation in natural tasks. *Philos Trans R Soc Lond B Biol Sci, 368*(1628), 20130066. doi: 10.1098/rstb.2013.0066

Thura, D., Hadj-Bouziane, F., Meunier, M., & Boussaoud, D. (2008). Hand position modulates saccadic activity in the frontal eye field. *Behav Brain Res, 186*(1), 148-153. doi: S0166-0.1016/j.bbr.2007.07.035

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychol Rev, 113*(4), 766-786. doi: 2006-12689-003

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *J Exp Psychol Hum Percept Perform, 24*(3), 830-846.

Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition, 8*(6), 769-800.

Tucker, M., & Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychologica, 116*(2), 185-203.

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Q J Exp Psychol (Hove), 59*(11), 1931-1949. doi: 10.1080/17470210500416342

Underwood, G., Foulsham, T., van Loon, E., Humphreys, L., & Bloyce, J. (2006). Eye movements during scene inspection: A test of the saliency map hypothesis. *European Journal of Cognitive Psychology, 18*(03), 321-342.

van der Linden, L., Mathot, S., & Vitu, F. (2015). The role of object affordances and center of gravity in eye movements toward isolated daily-life objects. *J Vis, 15*(5), 8. doi: 10.1167/15.5.8

Vishwanath, D., & Kowler, E. (2003). Localization of shapes: eye movements and perception compared. *Vision Res, 43*(15), 1637-1653.

Witt, J. K. (2011). Action's Effect on Perception. *Current Directions in Psychological Science, 20*(3), 201-206. doi: 10.1177/0963721411408770

## Figure Captions

**Figure 1.** Trial structure and exemplar stimuli. The figure shows the timeline of three exemplar trials in which a graspable object (Rubik's cube), an ungraspable object (coach), and a tool (peeler) were shown.

**Figure 2. Schematic representation of the FROA methodology.** The figure shows the computational steps carried out to compute the absolute and chance overlap percentages (AOP and COP, respectively) based on the participants' fixation gaze data and the visual saliency model(s) for an exemplar stimulus (peeler).

**Figure 3.** Results of the Model Matching Dissimilarity analysis. The figure shows the MMD values as a function of object Category (Tool, Graspable, and Ungraspable) and Body Posture (Unconstrained vs. Constrained). * indicates the significant effect of the body posture modulation at the Newman-Keuls's post-hoc test for the Body Posture by object Category interaction. # indicates the significance of the same effect at the post-hoc ANOVA on the corresponding difference scores. † indicates significant different body posture effects as compared to the Tool category. Error bars indicate *SEM*.

**Figure 4**. Distribution of fixation data for tools. The figure shows the normalized percentage of fixations (norm% values) as a function of Body Posture (Unconstrained vs. Constrained) and Tool Part (Functional, NFF; GBVS-Salient, $NFS_{GBVS}$; and AWS-Salient, $NFS_{AWS}$) both for each 500 ms-long time bin (A) and for each of the first five fixations (C). Panels B and D show the corresponding norm% values averaged across time bins and fixations, respectively. * indicates significant differences at the Newman-Keuls's post-hoc test for the Tool Part by Body Posture interaction.