Flusberg, S. J., Toskos Dils, A., & Boroditsky, L. (2010). Motor affordances in object perception.
Proceedings of the 32nd annual meeting of the Cognitive Science Society, Portland.

Motor Affordances in Object Perception

Stephen J. Flusberg (sflus@stanford.edu)
Alexia Toskos Dils (atoskos@stanford.edu)
Lera Boroditsky (lera@stanford.edu)
Stanford University, Department of Psychology
Jordan Hall, 450 Serra Mall, Building 420, Stanford, CA 94305 USA

Abstract

Recently, researchers have suggested that when we see an object we automatically represent how that object affords action (Tucker & Ellis, 2001). However, the precise nature of this representation remains unclear: is it a specific motor plan or a more abstract response code? Furthermore, do action representations actually influence *what* we perceive? In Experiment 1, participants responded to an image of an object and then made a laterality judgment about an image of a hand. Hand identification was fastest when the hand corresponded to both the orientation and grasp type of the object, suggesting that affordances are represented as specific action plans. In Experiment 2, participants saw an image of a hand before interpreting an ambiguous object drawing. Responses were biased towards the interpretation that was congruent with the grasp type of the hand prime. Together, these results suggest that action representations play a critical role in object perception.

Keywords: object perception; motor affordances

Background

Traditional approaches to visual perception have assumed that the primary goal of the visual system is to construct a detailed internal picture of the external world based on a noisy retinal image (e.g. Marr, 1980). Recently, however, there has been a growing appreciation for the possibility that visual perception may be equally concerned with how we move around and act in our environment (Milner & Goodale, 1995). The idea that vision and action are intimately linked can be traced to the ecological psychology of James Gibson (1979), who argued that organisms see the world in terms of how it affords *action*. While Gibson eschewed the notion of mental representation, contemporary scholars have suggested that visual perception may be at least partially characterized by a mental representation of the *affordances* in the environment (Tucker & Ellis, 1998). For example, seeing the coffee mug on the desk before me might involve mentally representing how I could reach out and grasp it in order to drink from it. However, the precise nature of these affordance representations and how they relate to object perception remains unclear.

Several researchers have suggested that affordances are represented as action plans in the motor systems of the brain (e.g. Tucker & Ellis, 1998; Chao & Martin, 2000). Tucker and Ellis conducted a series of studies to test whether people automatically generate a motor representation in response to the visual presentation of an object, even when there is no intention to act on the object (e.g. Tucker & Ellis, 1998, 2001). In one experiment, participants had to make a left or right-handed button press to indicate whether an image of an object on the screen was upright or inverted. The objects were chosen to have clear right or left-handed affordance (e.g. a frying pan with a handle oriented to the left affords a left-handed grasp). Participants responded faster and made fewer errors when their responding hand was congruent with the (task-irrelevant) affordance of the object on the screen. Similar stimulus/response compatibility (SRC) effects have been obtained for other types of *micro-affordances*, such as *grasp type* (Tucker & Ellis, 2001) and *wrist orientation* (Tucker & Ellis, 1998).

Others have argued that the motor representations activated when we observe an object form an integral part of our perception of the object (e.g. Helbig, Graf, & Kiefer, 2006). For example, Helbig et al. (2006) presented participants with images of two objects in quick succession. Participants were more accurate in naming the second object when similar actions were required to make use of two objects (e.g. pliers and a nutcracker). These findings raise important questions regarding, (1) the level at and specificity with which affordance information is represented, and (2) the potential causal role this information plays in the process of object perception.

For instance, several researchers have argued that the SRC effects obtained by Tucker and Ellis actually reflect abstract response coding rather than specific motor plans (e.g. Anderson, Yamagishi, & Karavia, 2002). Indeed, even Tucker and Ellis (2001) suggested that these effects could not be explained by appealing exclusively to the neural systems responsible for the on-line control of actions because they were obtained using both images of objects as well as real-world objects that were out of reach of participants. Rather, affordance representations might be more abstract, specifying, for example, the general class of hand shape required to interact with an object rather than precise motor parameters. Further, while the results of Helbig et al. (2006) are consistent with motor affordance representations contributing to object perception, the data could also be explained by the fact that objects that are used in similar ways are typically similar to one another in other ways as well (e.g. semantically).

In this paper we describe two experiments designed to address these issues. First, we wanted to know whether motor affordances might be represented as specific action plans for interacting with an object rather than as abstract response codes. To this end, in Experiment 1 we made use of a dependent measure that is known to draw on very specific manual action representations, namely the hand identification task (Parsons, 1987). Previous work has shown that the time it takes to identify whether an image is of a left or right hand is directly proportional to the time it would take, and how difficult it would be, to rotate your own hand into that position (Parsons, 1987). In our study, participants first made a response towards an image of an object that afforded a particular grasp type and wrist orientation. They then saw an image of a hand and had to indicate whether it was a right or left hand. If seeing an object leads to the activation of a specific manual action representation, participants should be faster to respond to an image of a hand that matches that object on grasp type and wrist orientation.

Second, we investigated the possibility that action representations actually contribute to the perceptual representation of objects. In Experiment 2, participants were first primed with an image of a hand depicting a specific grasp type. They then saw a drawing of an ambiguous object and had to indicate what they thought it was. The object could be interpreted as affording a power grasp (e.g. a football) or a precision grasp (e.g. a coffee bean). Responses were biased towards the interpretation that was congruent with the primed hand. We also included a control condition designed to rule out task demand and memory-based explanations of our findings. Together, these results suggest that action representations play a critical role in object perception.

Experiment 1

What motor information becomes activated when we look at an object? Previous research suggests that abstract response codes representing individual micro-affordances such as *grasp type* or *wrist orientation* become activated during object perception (Tucker & Ellis, 2001). Experiment 1 makes use of a novel application of the hand identification task in order to test the specificity of motor representations activated during object perception. Participants first made judgments on an image of an object that afforded a particular grasp type (power, precision, or no grasp affordance) at a particular wrist orientation (upright or inverted). Then they made laterality judgments on an image of a hand that was configured in a particular grasp type (power or precision) at a particular orientation (upright or inverted). To the extent that viewing objects activates specific manual motor plans selective for both grasp type and wrist orientation, laterality judgments should be fastest when *both* of the micro-affordances manipulated align between the images of the object and hand.

The degree to which various micro-affordances are activated during object perception might also depend on current goals (Bekkering & Neggers, 2002). Experiment 1 was designed to test whether viewing objects activates motor plans more strongly when participants make grasp-related compared with grasp-unrelated judgments about the objects. Similar reaction time profiles between these conditions would suggest that motor affordances are activated automatically during object perception, regardless of current task goals. Conversely, differences in reaction time profiles between the grasp-related and grasp-unrelated conditions would suggest that the current task goals do affect the kind of motor representation activated.

Methods

Participants Sixty-eight right-handed individuals from the Stanford community were recruited to participate in this study in exchange for payment or class credit.

Stimuli *Object Images*: Objects used in Experiment 1 varied on two dimensions: required *grasp type* (power, precision, or none) and required *wrist orientation* (upright or inverted). The dimensions were fully crossed within-subjects to produce 6 different object types. Two power grasp objects (flashlight and glass), two precision grasp objects (pushpin and tweezers), and four objects with no grasp affordance (desk, bookcase, grandfather clock, and sofa) populated the object categories. Hence, participants saw 16 unique images, each of which was repeated 32 times for a total of 512 object presentations.

The object stimuli were designed to afford right-handed responses because we recruited exclusively right-handed participants. The upright version of each object faced upward and to the left so as to afford an upright right-handed grasp on the part of the observer. The upright version of each object was rotated 90 degrees counter-clockwise in order to create the inverted version, which faced down and to the left. Pilot testing confirmed that right-handed individuals most often reached for real-world objects in both the upright and inverted orientations with their right hands.

Hand Images: The hand images varied on three dimensions: *grasp type* (power or precision), *wrist orientation* (upright or inverted), and *laterality* (left or right). The dimensions were fully crossed within-subjects to produce 8 different hand types. Four images of hands producing a power grasp and four images of hands producing a precision grasp were used to populate each of the hand categories. Hence, participants saw 32 unique hand images, each of which was repeated 16 times for a total of 512 hand presentations. The upright and inverted right and left hand images were generated using the same process described above for the upright and inverted object images.

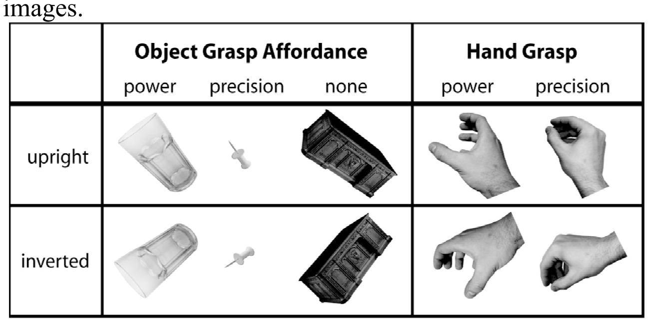

Figure 1: Sample stimuli from Experiment 1

Procedure Each trial in Experiment 1 had two parts. Participants first responded to a picture of an object and then to a picture of a hand. Participants were randomly assigned to one of two conditions: an *Orthogonal Judgment Condition* and a *Non-orthogonal judgment condition*. In the Non-orthogonal Condition, participants made a grasp-related judgment in response to each object ("Can you pick it up with one hand?"). In the Orthogonal Condition, participants made a grasp-unrelated judgment in response to each object ("Is it smaller than a shoebox?"). In both conditions, participants pressed the "j" key with their right index finger to enter a "yes" response, and the "f" key with their left index finger to enter a "no" response. Participants were told to respond as quickly and accurately as possible. Each object in Experiment 1 was preceded by a 500 ms fixation period. The image remained on the screen until the participant responded or the 10-second deadline expired, at which point the trial advanced to the hand portion.

Each hand image was preceded by a 500ms fixation period. Participants pressed the "j" key with their right index finger for pictures of right hands, and pressed "f" with their left index finger for pictures of left hands. The experiment advanced when the participant entered a response or at the end of the 10-second deadline. A black screen appeared for 750 ms to mark the end of each trial. The 16 unique object images were fully crossed with the 32 unique hand images to generate 512 unique experiment trials, each of which participants saw only once.

Results

The data from eight participants were removed because they did not contribute to all cells in the analysis or they had extremely high error rates or reaction times.

Trials analyzed: Only trials in which participants made correct responses to both the object and hand images were analyzed, resulting in the exclusion of 13.4% of trials. Any response times faster than 200 ms. or slower than 5000 ms. were omitted from all analyses, resulting in the removal of 1.4% of remaining trials across conditions. Finally, the stimuli used in Experiment 1 were designed to elicit right-hand affordances from right-handed individuals. As a result, only images of right hands were analyzed.

Coding: In 'Orientation Match' trials the orientation of the object was identical to that of the subsequent hand (collapsing across upright and inverted images). In 'Orientation Mismatch' trials the orientation of the object differed by 90 degrees in angular rotation from that of the subsequent hand.

RT Analyses: Figure 2 illustrates the mean pairwise RT differences (Orientation Match – Orientation Mismatch) across all levels of Object Affordance (power, precision, none) and Hand Stimulus (power, precision). Negative difference scores suggest a match advantage with respect to orientation. Positive difference score suggest a mismatch advantage with respect to orientation. The difference scores were submitted to a 3 (Object Affordance: power, precision, none) x 2 (Hand Stimulus: power, precision) repeated measures ANOVA. The analysis produced no main effects of object affordance ($F(2,57)=1.53$, *ns*) or hand stimulus ($F(1,58)=0.436$, *ns*), but a reliable quadratic interaction between the two variables ($F(1,58)=13.14$, $p< 0.001$).

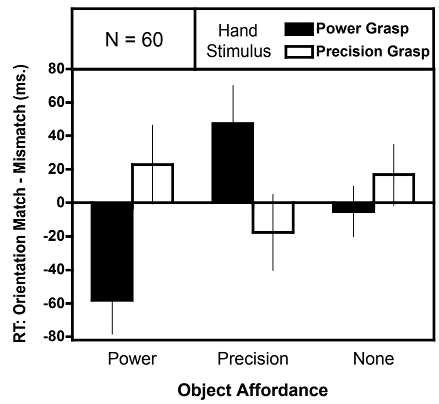

Figure 2: Differences in reaction time (Orientation Match – Mismatch) to the hand stimulus in Experiment 1. Error bars reflect the SE of the mean for each cell.

Participants showed a reliable match advantage to images of hands in a power grasp after having seen an object affording a power grasp (M=-58 ms, SD=158) compared to having seen an object with no manual action affordance (M=-5 ms, SD=116) ($t(59)=-2.31$, $p<0.05$). Conversely, participants showed a reliable mismatch advantage to images of hands in a power grasp after having seen an object affording a precision grasp (M=47 ms, SD=177) compared to having seen an object with no manual action affordance (M=-5 ms, SD=116) ($t(59)=2.05$, $p<0.05$). Images of hands in a precision grasp showed analogous trends, but none of the relevant comparisons reached significance (all $p>0.2$). This may be due to the fact that these hands were harder to correctly identify and thus they produced more errors and more variance in RT compared to grasp hands. The effect of orientation for hands in a precision grasp did, however, differ from the effect of orientation for hands in a power grasp both when the preceding object required a power grasp (M=22.80, SD= 23.51) ($t(59)=-2.58$, $p< .05$) and when the preceding object required a precision grasp (M=-17.54, SD=22.63) ($t(59)= 2.05$, $p< .05$).

The data appear to follow a Mexican hat distribution (Muller et al., 2005), such that reaction times *increase* when the object affordance only somewhat overlaps with the hand stimulus and *decrease* when the two overlap entirely (relative to trials where the preceding object had no grasp affordance). To further test for such a distribution, trials were binned into five similarity-based categories (Figure 3). For each of the bins, the object affordance on a given trial relative to the subsequent hand stimulus was either: (1) same orientation and grasp type, (2) same grasp type only, (3) same orientation only, (4) different orientation and grasp type, or (5) had no grasp affordance. A repeated measures

ANOVA yielded reliable quadratic ($F(1,58)=8.04$, $p<0.01$) and cubic ($F(1,58)=6.05$, $p<0.02$) effects of similarity.

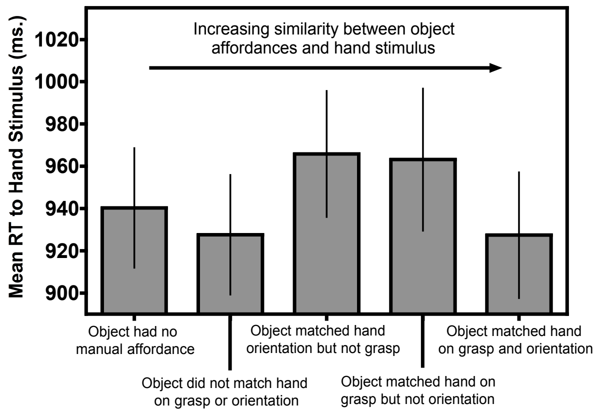

Figure 3: RT to hand stimuli in Experiment 1, binned by how similar the hand was to the set of manual affordances in the previous object

Finally, there was no main effect of the between-subjects condition (Task Goals: grasp-related, grasp-unrelated) ($F(1,58)=1.50$, $p>0.20$), nor did condition interact with the quadratic Object Affordance by Hand Stimulus interaction ($F(1,58)=2.81$, $p>0.10$) or the effect of similarity ($F(1,58)=1.76$, $p>0.15$). As a result, all analyses described above were run on the combined data from the two conditions.

Discussion

In this experiment we asked whether people generate a specific motor plan for interacting with an object when they see that object. RTs for the hand laterality judgment were fastest when the hand corresponded in both grasp type and wrist orientation to the previous object. This suggests that when we look at an object we represent the specific motor parameters necessary for interacting with that object. In other words, when we see a drinking glass we actually simulate reaching out and grasping it.

Interestingly, RTs for the laterality judgment were slowest when the object and hand corresponded in just one micro-affordance dimension (i.e. either grasp type *or* wrist orientation). Researchers have argued that similar reaction time profiles in classic vision and attention tasks suggest an underlying surround inhibition mechanism (Muller et al., 2005; Roeber, Wong, & Freeman, 2008), where activating a particular representation suppresses highly similar but not distantly similar representations. Importantly, motor cortex is believed to have the kind of connectivity that would support surround inhibition (Lukashin & Georgopoulos, 1993; Sohn & Hallett, 2004). Thus it is plausible that viewing an object in the present study activates a highly specific action representation, which in turn spreads inhibition to highly similar but not distantly similar action representations. These patterns of activation and inhibition would result in slower RTs to trials in which the images of the object and hand differ on one but not all dimensions, which is precisely what was found in Experiment 1.

Such connectivity further predicts a full Mexican hat-like distribution of response times such that the representations in similarity space just beyond the inhibited surround should see facilitation that tapers off as distance increases (Muller et al., 2005). That is, responses to hands preceded by objects that afford the wrong grasp in all dimensions should be faster than those preceded by objects that afford no grasp at all. The cubic effect of similarity found in Experiment 1 is driven by that very difference, suggesting a Mexican hat response time profile (Muller et al., 2005). Studies better designed to test for such a pattern are currently underway. As it stands, the pattern observed in Experiment 1 is consistent with the kind of connectivity believed to exist in motor cortex. Furthermore, the hand identification task used in this study was selected precisely because it is believed to be supported by specific motor regions. As a result, the present findings are consistent with the hypothesis that object perception activates highly specific action representations in the motor system and does so in a manner similar to the act of grasping itself.

Finally, varying participants' task goals when viewing the object had no significant effect on these results. Whether participants made a grasp-related ("Can you pick it up?") or grasp-unrelated ("Is it smaller than a shoebox?") judgment towards the object, the same affordance information appears to have been represented. This supports the original findings of Tucker and Ellis (1998), who argued that affordance information is represented irrespective of the intentions of the observer. However, other researchers have found effects of intentions on affordance representation (e.g. Bekkering & Neggers, 2002), and the effects in the present study tended to be more robust in the task-related than the task-unrelated condition, which suggests that more research is called for on this issue.

Experiment 1 supports the idea that motor affordances are represented as specific action plans in the motor system regardless of task goals. However, it is unclear how this action representation relates to our ability to actually perceive the object. We turn to this issue in Experiment 2.

Experiment 2

Does action representation contribute to object perception? One possibility is that the visual and motor aspects of object perception are fairly independent: extracting the visual features of an object occurs in one processing stream while extracting the affordance information relevant for action occurs in a different processing stream (Milner & Goodale, 1995). Alternatively, visual and motor processes may be more interdependent, and currently activated action representations might play a causal role in visual object processing. Experiment 2 was designed to test the latter possibility by priming participants with a specific manual action to see if it would affect their interpretation of an ambiguous object drawing. We also ran a control condition where we presented the ambiguous object image first in order to control for task demand or memory-based explanations for the data.

Methods

Participants 245 individuals from Amazon's Mechanical Turk website participated in exchange for payment.

Stimuli & Procedure The stimuli for this experiment included four photographs of hands taken from Experiment 1 and an ambiguous object line drawing created by the authors. The four hand photographs showed either left or right hands in either a power or precision grasp. Pilot testing suggested that the ambiguous object drawing could be interpreted as an object that afforded a power grasp (e.g. *football*) or as an object that afforded a precision grasp (e.g. *coffee bean*).

In the *experimental condition*, one of the four hand images was randomly selected for each participant and displayed on the screen for three seconds. After this, the ambiguous object drawing was displayed for three seconds. Then, participants were asked to name the object in the line drawing that they had just seen and to identify whether the hand they had seen was a left or right hand. The only difference in the procedure for the *control condition* was that participants were shown the ambiguous object image first and hand image second. Thus participants were not primed with an action representation prior to viewing the ambiguous object, but they saw the same two images prior to making their object interpretation response.

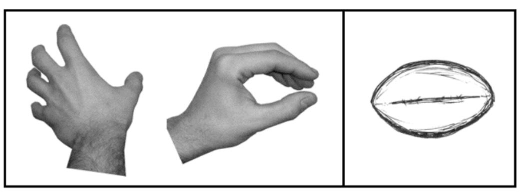

Figure 4: On the left, images of the left hand stimuli used in the experiment displaying precision and power grasps. On the right, the ambiguous object drawing.

Results

The data from 19 participants were removed because they failed to respond to the test questions appropriately (N=8) or because they took the survey more than once (N=11). We then coded the object interpretation responses for the remaining participants in terms of what sort of grasp would be afforded by the perceived object. We used the following coding scheme: Power (e.g. *football* or *coconut*), Precision (e.g. *coffee bean* or *nut*), and None (e.g. *lips* or any response that listed more than one interpretation).

Experimental Condition: For our analyses we collapsed across left and right hand prime stimuli and excluded object interpretation responses coded as *none*. A 2 (hand prime stimulus: power grasp vs. precision grasp) X 2 (perceived object affordance: power vs. precision) chi-square test of independence showed a significant relationship between hand prime stimulus and perceived object affordance, $\chi^2 = 7.04$, $p<0.01$. Participants primed with an image of a power grasp hand were more likely to interpret the ambiguous image as an object affording a power grasp while participants primed with a precision grasp hand were more likely to interpret the ambiguous image as an object affording a precision grasp.

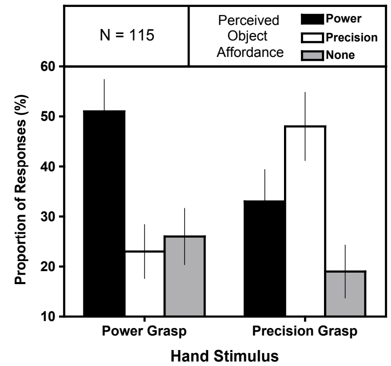

Figure 5: Experiment 2, *Experimental Condition*: Proportion of ambiguous object interpretations coded for perceived object affordance. Error bars are the SE of the proportion.

Control Condition: A 2 (hand stimulus: power grasp vs. precision grasp) X 2 (perceived object affordance: power vs. precision) chi-square test of independence showed no relationship between hand stimulus and perceived object affordance, $\chi^2 = 0.74$, $p>0.38$.

Interaction Analysis: A 2 (interpretation: congruent vs. incongruent) X 2 (condition: experimental vs. control) interaction analysis showed that hand stimuli only affected ambiguous object interpretations in the experimental condition, $\chi^2 = 4.05$, $p<0.05$.

Discussion

In this experiment we asked whether currently active motor representations would influence what participants saw when they looked at an ambiguous object. We found that when participants were primed with a hand displaying a power grasp they were more likely to interpret an ambiguous image as an object that afforded a power grasp (e.g. football). Conversely, when they were primed with a hand displaying a precision grasp, they were more likely to interpret the image as an object that afforded a precision grasp (e.g. coffee bean). This finding suggests that action representations can play a causal role in the process of object perception.

That said, there are a number of possible alternative explanations for these data. First, because of the simple design of this experiment, participants may have simply figured out what we wanted from them and tried to give it to us. The results of the control condition suggest that this is unlikely, however. In that condition, participants also saw both the ambiguous object and the hand stimulus prior to giving their object interpretation response, the only difference being they saw the ambiguous object first. If the results from the experimental condition were due to demand characteristics, we would expect to find the same pattern of results here. However, in the control condition there were no such effects. This also helps rule out the possibility that the results of the experimental condition were due to associations in memory rather than the online effects of action representation on perception.

Finally, because we used purely visual stimuli in this experiment, it is possible that our results reflect visual priming rather than motor priming. While prior research has demonstrated that visually processing images of hands typically involves activating motor representations of one's own hand (Parsons, 1987), it is difficult to rule out visual priming as an explanatory mechanism at the present time. Research currently underway in our lab is moving away from visual prime images and towards actual motor movements as priming stimuli. Moreover, we are developing additional controls that include reversible images that do not afford grasping in order to rule out alternative mechanisms such as altered scanpaths or attentional patterns.

General Discussion

In this paper we explored the role that action representation plays in visual object processing. In Experiment 1 we took a bottom-up approach, asking whether we generate a specific motor plan or a more abstract response code when we observe an object with a particular set of manual affordances. Participants made a judgment about an image of an object that afforded a particular grasp type and wrist orientation. They then made a laterality judgment about an image of a hand displaying a particular grasp type and wrist orientation. RTs for the laterality judgment were fastest when the hand corresponded in both grasp type and wrist orientation with the previous object. This suggests that when we look at an object we represent the specific motor parameters necessary for interacting with that object within the motor systems of the brain.

Intriguingly, RTs for the laterality judgment were slowest when the object and hand corresponded in just one micro-affordance dimension (i.e. either grasp type *or* wrist orientation). This "Mexican hat" response time function has been found by other researchers studying motor representation in the brain (Loach, Frischen, Bruce, & Tsotsos, 2008; Lukashin & Georgopoulos, 1993; Sohn & Hallett, 2004), providing further support for the idea that affordances are represented as specific action plans in the motor system.

In Experiment 2 we took a top-down approach, asking whether activating a particular manual action representation would influence the perception of an ambiguous object image. The results suggest that action representations can play a causal role in the process of object perception.

All together, the results of these experiments suggest that action representation plays a crucial role in visual object processing. As we look around the world we are not merely constructing an internal picture of what's out there, we are also preparing to act and behave on what's before us. Furthermore, our current action state affects how we process the *what* that is out there.

Acknowledgments

The authors would like to thank the members of Cognation for their helpful feedback. This research was supported by an NSF Career Award Grant given to Lera Boroditsky.

References

Anderson, S. J., Yamagishi, N., & Karavia, V. (2002). Attentional processes link perception and action. *Proceedings of the Royal Society of London B, 269,* 1225-1232.

Bekkering, H. & Neggers, F. W. (2002). Visual search is modulated by action intentions. *Psychological Science, 13,* 370-374.

Chao, L. L. & Martin, A. (2000) Representation of manipulable man-made objects in the dorsal stream. *Neuroimage, 12,* 478-484.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Lawrence Earlbaum: Hillsdale, NJ.

Helbig, H. B., Graf, M., & Kiefer, M. (2006). The role of action representations in visual object recognition, *Experimental Brain Research, 107*(2), 221-228.

Loach, D., Frischen, A., Bruce, N., & Tsotsos, J. (2008). An attentional mechanism for selecting appropriate actions afforded by graspable objects. *Psychological Science, 19*, 1253-1257.

Lukashin, A. V., & Georgopoulos, A. P. (1993). A dynamical neural network model for motor cortical activity during movement: population coding of movement trajectories. *Biological Cybernetics, 69*, 517 – 524.

Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.

Milner, D. & Goodale, M. (1995). *The Visual Brain in Action*. Oxford: Oxford University Press.

Muller, N. G., Mollenhauer, M., Rosler, A., & Kleinschmidt, A. (2005). The attentional field has a Mexican hat distribution. *Vision Research, 45*, 1129-1137.

Parsons, L. M. (1987). Imagined spatial transformation of one's body. *Journal of Experimental Psychology: General, 19,* 178-241.

Roeber, U., Wong, E. M., & Freeman, A. (2008). Cross-orientation interactions in human vision. *Journal of Vision, 8*, 1-11.

Tucker, M. & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance*. *24*(3), 830-846.

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