(2001) Proceedings of the Twenty-third Annual Meeting of the Cognitive Science Society, (pp.867-872), Erlbaum: Mawhah, NJ

Motor Representations In Memory And Mental Models: The Embodied Zork

Daniel C. Richardson (dcr18@cornell.edu)
Michael J. Spivey (mjs41@cornell.edu)
Jamie Cheung (jmc56@cornell.edu)
Cornell University, Department of Psychology
Uris Hall, Ithaca, NY 14853

Abstract

A variety of experimental results have suggested that motor systems can participate in what were thought to be purely perceptual tasks. Extending previous work in the stimulus-response compatibility paradigm, we show that a representation of a visual stimulus accessed from memory can activate potential motor interactions. Other research has shown that mental images and mental models can have analogue or spatial characteristics. In our second experiment, we present evidence that such representations, generated purely from linguistic descriptions, can also activate motor affordances. Results are discussed within the context of the embodiment of *conceptual*, as well as perceptual, processes

Introduction

One can often observe people who, having lost the set of keys they held a moment ago, mentally retrace their steps, miming their previous interactions with objects and places that might now be a hiding place. Similarly, when imagining the rearrangement of a room, people often move their hands as if picking up and moving furniture. The gestures that embellish discourse are also made by people who are on the telephone, and even by the congenitally blind (Iverson & Goldin-Meadow, 1998).

Hand waving examples such as these offer glimpses of the relationship between 'higher' cognitive functions such as language and imagination, and the seemingly more mundane perceptual and motor systems that carry out our daily chores. Increasingly, it is suggested that these motor systems have an important contribution to cognition.

This zeitgeist of 'embodiment' has been described as 'a seismic event … taking place in cognitive science' (Newton, 1998) and a 'perennial recycling of behaviourist ideology' (Pylyshyn, 2000). In general, proponents of an embodied perspective reject the idea of cognition as wholly the processing of abstract, or amodal, symbolic representations. They emphasise the ways in which the representational and processing burden of cognition can be offloaded on the external world and the motor and perceptual systems that interact with it (Barsalou, 1999a; Clark, 1997; Lakoff & Johnson, 1999)

There is a growing weight of behavioural and brain imaging work that implicates motor systems in perceptual judgements. Observers often interpret visual stimuli in terms of the physically plausible motions it would take to produce them (eg Shiffar & Freyd, 1993). Bargh, Chen and Burrows (1995) present intriguing evidence that even sophisticated social constructs are imbued with motoric representations, to the degree that activating a stereotype will automatically cause motor behaviour typical of that social group.

In some cases, brain imaging techniques have shown the direct involvement of motor areas in 'motor perception' (Stevens, Fonlupt, Shiffrar & Decety, 2000). The existence of 'mirror neurons' also indicates that visual and motor systems share neural circuitry (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996). Behavioural experiments by Wohlschläger and Wohlschläger (1998) demonstrated that when subjects mentally rotated a 3D object, performance was slowed if the response used a rotational motor action that was in the opposite direction to the mental rotation. De'Sperati and Stucci (2000) argued that the motor system acts to simulate rotations: in their studies, subjects could more easily judge the rotation of a screwdriver if it was pictured in an orientation easily graspable by their dominat hand. This work is supported by studies that find activation of motor areas during mental rotation tasks (eg. Richter, Somorjai, Summers, Jarmasz, Menon, Gati, Georgopoulos, Tegeler, Ugurbil & Kim, 2000)

Recent research in the stimulus-response compatibility paradigm has found further evidence of the intrusion (or participation) of task-irrelevant motor representations during a perceptual judgement (e.g. Craighero, Fadiga, Rizzolatti & Umilta, 1998). In experiment 1 of Tucker and Ellis (1998) (henceforth T&E) subjects made an orientation judgement (right-side-up/upside-down) about pictures of household objects such as a coffee mug. Each object had an affordance - a handle - on the right or the left side. It was found that subjects were faster when they responded using the hand that was on the same side as the affordance. Later work (Ellis & Tucker, 2000) found a similar compatibility effect when subjects signalled a judgement about an object with a motor action (precision pinch/power grasp) that was appropriate or inappropriate for that object.

This work demonstrates a tight coupling between visual and motor systems: the perception of a graspable object immediately actives a potential motor interaction with that object, even though the affordance is irrelevant to the perceptual judgement. Yet is this just evidence of a rapid link, a transient information hook-up, between visual and motor systems, or is it indicative or a more long term relationship, whereby the representation of objects is not merely visual, or an amodal list features, but has a motor component that is just as much part of the object bundle.

In Barsalou's (1999a) Perceptual Symbol Systems theory, motor activations such as those revealed in T&E's work should become part of the long term 'simulator' or concept of an object. Klatzky, Pelligrino, McClosky and Lederman (1993). It has been shown that there are strong associations between object-action pairs (Klatzky, Pelligrino, McClosky & Lederman, 1993) and that object recognition and object grasping utilise overlapping neural networks (Faillenot, Toni, Decety, Gregoire & Jeannerod, 1997). It would be of further interest to show that simply activating a representation of an object causes activation of potential motor interactions. In the first experiment presented here, we extend the results of T&E to show an effect of response compatibility when subjects are recalling a visual image from short term memory.

A considerable body of research demonstrates that conceptual processing has some of the hallmarks of perceptual processing (cf. Barsalou, 1999a). For example, the 'scanning' of mental images mimics the time course of scanning a real image; if a certain location in a mental model is activated, the nearby locations are primed (Bower & Morrow, 1990). If, as the work above suggests, motor activation is often part of perceptual processing, then, we reasoned, perhaps it will be part of conceptual processing as well. The second experiment investigates the role of motor systems in a typical conceptual task – listening to a story.

Experiment 1

We have seen that the presentation of an object's image can activate a potential motor interaction that causally effects a motor response to that stimulus. In this experiment, we would like to see if affordances can be stored in, or reactivated by, short term visual memory.

Method

Subjects 40 undergraduate students of Cornell University participated in exchange for extra credit. All subjects were right handed

Stimuli We compiled 198 images of household objects by using a digital camera and searching public domain image databases. Each picture was in full colour, with a resolution of 500 x 400 pixels. 154 of these were filler images of objects with no obvious affordance, or with an affordance that could be accessed by both the left and right hand equally. We obtained 22 images of objects that had a clear affordance on one side; each of these each of these was mirror reversed. Some stimuli were taken, with kind permission, from Carlson-Radvansky, Covey and Lattanzi (1999). 40 sound files were recorded by an experimenter: the names of the 22 afforded objects, and the names of 18 objects that did not appear in the stimuli set.

Design Each subject was randomly assigned to a response mapping condition. In Left condition, subjects responded 'yes' by pressing the 'S' key and 'no' by the 'K' key. In the Right condition this mapping was reversed.

A schematic of each trial is given in Figure 1. Each trial began with the subjects being reminded of their response mapping. They then pressed the space bar to initiate the trial. Subjects were given a countdown from 3, and then saw 8 images in rapid succession. Each presentation lasted for 200ms. After a 1000ms pause, subjects heard the name of an object. They responded as quickly as possible whether the object was present in the set of 8. No feedback was given.

There were 40 trials. In 18 filler trials, the named object was not present. In the remaining 22 trials, an afforded object was present at a random location, and subsequently

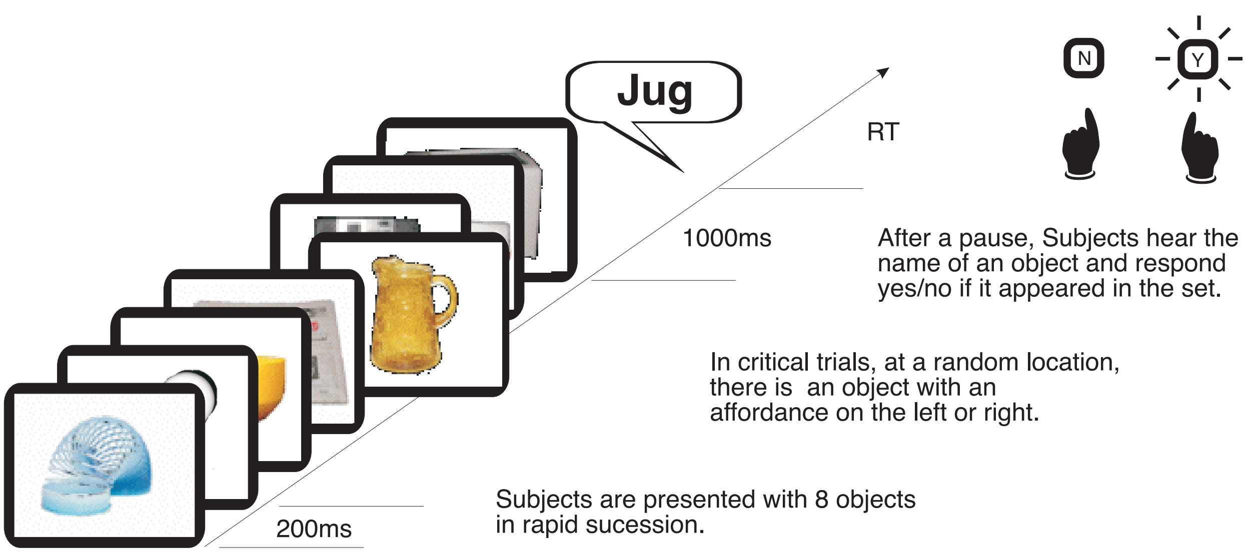

Figure 1: Schematic of Experiment 1.

named. Half of these objects had an affordance on the left and half on the right, counterbalanced across subjects. The order of the trials and was fully randomised.

Subjects were asked beforehand not to try and name the objects as they appeared. The use of very short presentation times was designed to discourage this verbal labelling strategy. After the experiment., subjects were debriefed and asked to what degree they were able to comply with these instructions.

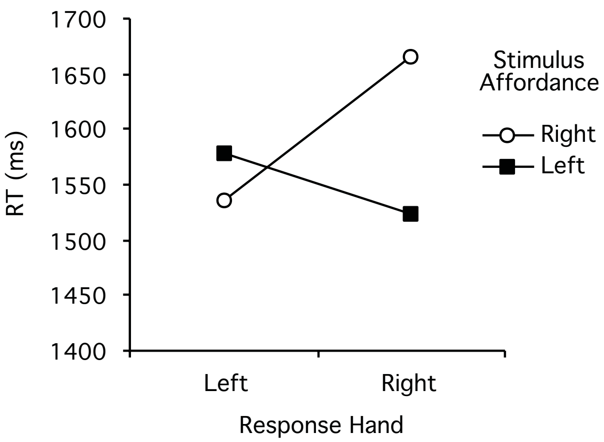

Figure 2 Mean RTs for correctly answered, critical trials, Exp.1

Results In the critical trials, the correct answer was always 'yes'. Accuracy on these trials was 74%. For the remaining analyses, we discarded trials with incorrect responses, and trials with an RT longer than 2.5 standard deviations from the mean (0.006% of the data). Figure 2 shows the trimmed data set.

Subjects making their 'yes' response with the left hand (1551ms) were marginally faster than the those using the right (1601ms), although this effect of response mapping did not approach significance ($F(1,38)=0.20$, $p>.6$). Responses were made slightly quicker when the named object had an affordance on the left (1556ms) versus the right (1598ms), but again this effect did not approach significance ($F(1,38)=1.72$, $p>.19$).

Yet as Figure 2 shows, there was a robust interaction between the stimulus and response conditions that was significant ($F(1,38)=8.22$, $p>.01$). When the hand making the response was on the same side as the affordance of the named object (compatible trials), the mean RT was 1611ms; when the response and object affordance were opposite sides (incompatible trials) the mean RT was almost 90ms faster at 1524ms.

Discussion

Like T&E, we have found an interaction between stimulus affordance orientation and response hand. Yet our results have some interesting differences. First, we have found an incompatibility effect, such that subjects are facilitated when making a judgement using the hand on the *opposite* side to the stimulus affordance. Second, our subjects have response times of about 1500ms, whereas T&E's subjects responded in roughly 700ms.

We suggest that it is primarily this second factor that helps explain the difference in the direction of compatibility. Work by Stoet and Hommel (1999; 2000) shows an interesting time course to the activation and binding of stimulus and response features. These authors use the framework of the Theory of Event Coding, which holds that perception and actions are coded in a common medium. They showed that up until a certain time, stimulus features can be activated such that they facilitate compatible or overlapping responses. However, when the features are activated for a longer time period, they can become bound into an "event file". Once bound, those features are less available for coding compatible responses, and hence an incompatibility effect is found.

Experiment 2

We have found evidence for an effect of compatibility between a motor response and the affordance of a visual stimulus under conditions where subjects are recalling the relevant object from memory over the course of several seconds. We have seen that there are interesting suggestions form other work that, (a) the direction of the effect hinges upon the time course of feature activation, and (b) these visuo-motor feature codes may be inherent to the representation of functional objects.

If it is the case the motor activation can occur with - perhaps be part of - the activation of object representations, then it should be possible to generate stimulus response compatibility effects from non-visual, verbal descriptions. The difficulty is how to make subjects imagine an afforded object in a particular orientation. Of course, if subjects heard, 'A jug with a handle on the left', then it could be argued that any spatial compatibility effects would be generated by the word 'left' rather than anything related to a motor component of the representation of 'jug.

We attempted to solve this problem by constructing rich scene descriptions. Subjects listened to these stories and then made a yes/no key press in response to a question. The critical trials contained sentences in which the location and orientation of an afforded object was implied by reference to other objects. As in experiment 1, for critical trials this question pertained to the afforded object, and the correct answer was 'yes'. This design allowed us to investigate whether the imaginary orientation of object - indirectly and verbally described - could interact with the hand used to make a key press.

Pilot work with these stories suggested that subjects made their responses between about 400 and 1800ms – roughly the spanning the RTs of T&E's subjects and our own in Experiment 1. Therefore, we decided to probe the time course of any feature activation by splitting the data at the median RT. We hypothesised that responses below this time would follow a S-R compatibility pattern similar to T&E's, whereas response over the median would have an incompatibility effect resembling Experiment 1.

Method

Subjects 110 right handed Cornell undergraduates participated in this experiment in exchange for course credit. None of the subjects had previously run in Experiment 1.

Stimuli 24 short scene descriptions and questions were written and recorded by the experimenter. The present tense was used throughout. Half of these stories were used as filler items, and the question related to the property of some object or person in the story. The other half of the stories were used as critical trials. Each of these included a description an object with an affordance, and specified the orientation of that object by reference to surrounding items.

First, there was a sentence or two conveying the background scene; in this case, a breakfast table. Two items, one on either side of the scene, were then described. We termed these the 'anchor' objects. In Figure 3 these are a bowl of cornflakes and an egg cup. Then, a third object was mentioned. This was the critical item, an object with an affordance, and was located between the two anchors. Then two phrases specified the orientation of the critical item. They linked a feature or affordance of the critical object with each of the anchors. In the Figure 3 example, the milk jug handle is next to the egg cup, and the spout is pointing towards the bowl. A sentence or two ended the scene description with some further background information. The question was some form of, 'In the center of the [scene] was there a [afforded object]?' The answer to this question on critical trials was, of course, 'yes'.

Figure 3 shows the structure of an example critical trial[1]. This structure allowed us to counterbalance two factors between subjects. Firstly, we varied the left/right positions of the two anchor objects, and hence the afforded object's orientation. Secondly, we switched the order of the two phrases linking afforded object features to anchor objects. This meant that between subjects, we could counterbalance whether a right sided or left sided object was referred to last. Hence any response biases could not be accounted for in terms of simple recency effects.

Design Each subject was randomly assigned to a response mapping condition, as in Experiment 1. Before each trial, subjects were reminded of the response keys to use. There were 24 trials. Each began with a short scene description, about 30 seconds in duration, played over a set of headphones. At the end of each description, subjects heard a one second tone and then a question concerning the previous information. They were instructed to give their response as quickly and as accurately as possible. Although subjects could take as long as they wanted, if their response time exceeded 5 seconds, that trial was not used in further analysis. We were primarily interested in how subjects represented and manipulated information in order to answer the question, not whether or not they could remember the description verbatim. The cut off point was set on the basis of pilot data to exclude trials in which subjects were struggling to remember details.

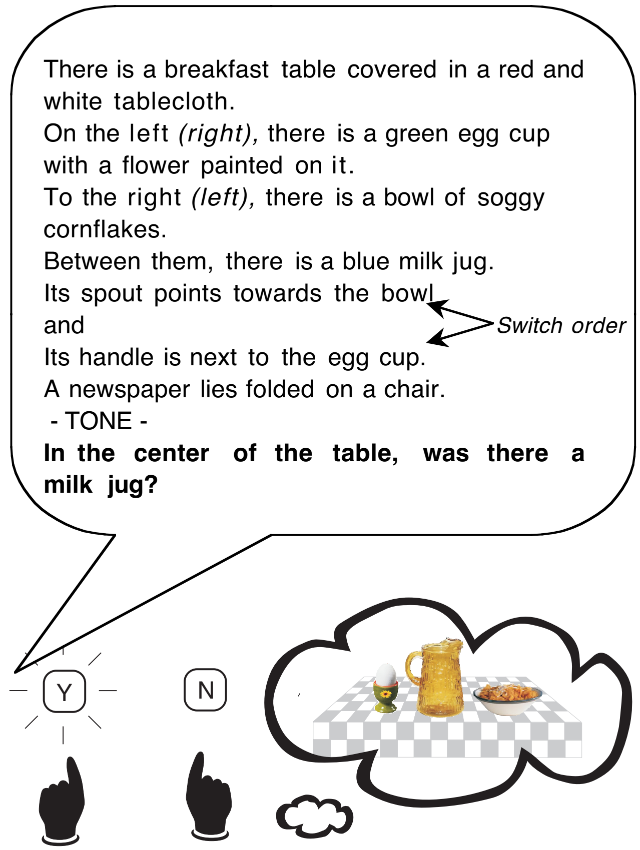

Figure 3: Schematic of a critical item, Exp. 2.

Results

The accuracy rate on critical trials was 84.5%. As in Experiment 1, only correct answers to critical trials were analysed. RTs more than 2.5 standard deviations from the mean were excluded from the analysis (4.8% of the data).

The mean RT was 1180ms. The subjects making right handed responses (1158ms) were slightly faster than those using the left hand (1201ms), but this effect did not approach significance ($F(1,108)=0.32$, $p>.5$). In addition, neither the effect of stimulus affordance ($F(1,108)=0.48$, $p>.4$), nor the hand x stimulus interaction ($F(1,108)=0.06$, $p>.8$) approached significance.

To test our hypothesis concerning the time course of stimulus response compatibility effects, we split the trials at the median RT (1020ms) into late and early groups. In order to carry out ANOVAs, we had to remove subjects who did not contribute to both cells. There remained 82 subjects in the early response condition, and 90 in the late. Their results are shown in Figure 4.

In both groups the main effects of response hand (early, $F(1,80)=2.33$, $p>.1$; late, $F(1,88)=1.12$, $p>.29$). and stimulus affordance (early, $F(1,80)=0.30$, $p>.5$; late, $F(1,88)=0.32$,

[1] A colleague, Reuben Saunders, remarked on the similarity between the present tense, second person prose of our scene descriptions, and the text-based computer role-playing game, Zork.

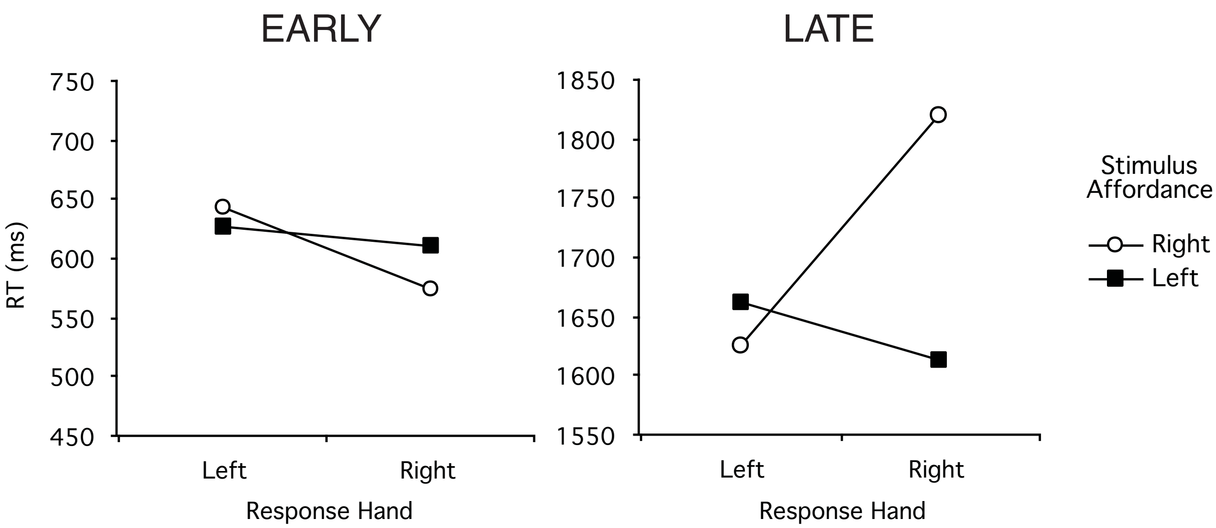

Figure 4: RTs of Experiment 2, split at the median of 1020ms.

p>0.18) did not approach significance. There was a significant interaction between hand and stimulus in the late group; $F(1,88)=4.38$, $p<.04$. This interaction was not significant in the early responses; $F(1,80)=0.30$, $p>.5$.

Thus, the significant hand X stimulus interaction in the late responses clearly shows an S-R *incompatibility* effect for responses that take place after 1020 ms. In contrast, the early responses show a numerically inverted, albeit nonsignficant, interaction - suggesting an S-R *compatibility* effect for responses that take place before 1021 ms. It is noteworthy that the magnitude of the S-R compatibility advantage for these early responses (20-30 ms) is numerically comparable to that found by T&E.

Finally, in order to get some quantitative indication of whether the interactions taking place in the two time periods are indeed different from one another, we conducted a three-way ANOVA that included early vs. late group as a factor. (Strictly speaking, it is improper treat this factor, which is derived from a dependent variable, as an independent variable. However, the test of a three-way interaction should be sensitive to the relative reaction times across conditions rather than the raw reaction times, and therefore should not be unfairly affected by this procedural irregularity.) As it was necessary to exclude participants who did not contribute to all cells of the design -- many participants were always fast or always slow -- the three-way ANOVA was conducted on the remaining 61 subjects. As predicted, a reliable three-way interaction was obtained where the early reaction times showed a pattern consistent with an S-R compatibility effect and the late reaction times showed a pattern consistent with an S-R incompatibility effect; $F1(1,59)=4.995$, $p<.05$.

Discussion

As hypothesized, the early and late responses show opposite stimulus-response compatibility effects, offering support for the feature activation integration model of Stoet and Hommel (1999). Moreover, we have shown that even in the prime 'disembodied' activity of language comprehension, subjects employ motor representations to construct a mental model, much as Stein's (1994) METATOTO robot created internal maps out of its motor interactions. Thus we have found empirical evidence in support of Bryant's (1998) claim - 'the internal worlds we create do not form maps of external space per se, but of perceptual and behavioral affordances within space.'

General Discussion

Research within the stimulus-response compatibility paradigm has shown that there is a tight coupling between perception and action; indeed, their function is so intimate that it suggests a 'common coding' of perceptual and motor features (Hommel, Müsseler, Aschersleben & Prinz, 2001).

The current experiments reveal one way in which conceptual processes intersect with this tight perception-action arc. Object representations, whether they are memories of a visual stimulus, or part of a mental model generated from a linguistic description, contain motor representations. These results show how motor systems take can take part in 'higher' cognitive functions.

Previous work in our laboratory has shown how *oculo*motor systems participate in the comprehension of spatially extended narratives (Spivey, Tyler, Richardson & Young, 2000), and the spatial indexing of linguistic information (Richardson & Spivey, 2000). Gold and Shadlen (2000) found that in the monkey cortex, competing patterns of activation in populations of the *motor control* neurons will themselves instantiate a 'decision' to saccade. These cases of motor activation occurring as part of a cognitive process are complimented by examples of linguistic representations intruding upon motor processes. Gentilucci, Benuzzi, Bertolani, Daprati, and Gangitano (2000) showed how automatically activated linguistic representations can intrude upon motor processes. They found that words such as ‘near’ and ‘far’ taped on a small wooden bar systematically modulated the kinematics of subjects’ reaching behaviour. Moreover, the grammatical class of words, adjectives or adverbs, differentially affected motor control. The results presented in this paper show that the motor system can be modulated not just by spatially extended words, but also by descriptions of afforded objects.

Barsalou (1999b) observed that language is often framed as a means for archiving knowledge. In contrast, he argued for a conception of language comprehension as a ‘preparation for situated action’. The involvement of motor representations in memory and mental models we have shown here suggests that language is aptly embodied for this function.

Acknowledgements

The authors wish to thank Bernhard Hommel and Mike Tucker for encouraging and insightful correspondence. Supported by a Sloan Foundation Fellowship in Neuroscience to MJS, and a Sage Fellowship to DCR

References

Bargh, J. A., Chen, M., Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71(2),* 230-244

Barsalou, L.W., (1999a). Perceptual symbol systems. *Behavioral and Brain Sciences, 22(4),* 577-660.

Barsalou, L.W., (1999b). Language comprehension: archival memory or preparation for situated action? *Discourse Processes, 28(1).* 61-80.

Bower, G.H., & Morrow, D.G. (1990). Mental models in narrative comprehension. *Science, 247,* 44-48.

Bryant, D.J., (1998). Human spatial concepts reflect regularities of the physical world and human body. In Oliver & Gapp (Eds.) *Representation and processing of Spatial Expressions*. LEA: London.

Carlson-Radvansky, L.A., Covey, E.S., & Lattanzi, K.M., (1999). What effects on "where": Functional influences on spatial relations. *Psychological Science,10(6),*516-521.

Clark, A. (1997). *Being there: Putting brain, body, and the world together again.* MIT press: Cambridge, Mass.

Craighero, Fadiga, Rizzolatti, Umita. (1998) Visuomotor priming. *Visual Cognition, 5(1/1)*, 109-125.

de’Sperati, C., & Stucchi, N., (2000). Motor imagery and visual event recognition. *Experimental Brain Research, 133,* 273–278.

Ellis, R. & Tucker, M., (2000). Micro-affordance: The potentiation of components of action by seen objects. *British Journal of Psychology 91(4)*, 451-471.

Faillenot, I., Toni, I., Decety, J., Grégoire M., & Jeannerod, M., (1997). Visual pathways for object-orientated action and object recognition: functional anatomy with PET. *Cerebral Cortex*, *7(1),* 77-85.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G., (1996). Action recognition in the premotor cortex. Brain, Volume 119(2), 593-609.

Gentilucci, M., Benuzzi, F., Bertolani, L., Daprati, E., & Gangitano, M. (2000). Language and motor control. *Experimental Brain Research, 133(4),* 468-490.

Gold, J.I., & Shadlen, M.N., (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature, 404(6776*), 390-394

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding: a framwrok for perception and action planning. *Behavioral and Brain Sciences.* In press.

Iverson, J.M., & Goldin-Meadow, S., (1998). Why people gesture when they speak. *Nature, 396(6708*), 228.

Klatzky, R.L., Pellegrino, J.W., McClosky, B.P., & Lederman, S.J. (1993). Cognitive represenations of functional interactions with objects. *Memory and Cognition, 21(3),* 294-303.

Lakoff, G., & Johnson, M., (1999). *Philosophy in the flesh.* Bascis Books: New York, NY.

Newton, N. (1998). Review. Being there: putting brain, body and the world together again. *American Journal of Psychology,* 111(1).

Pylyshyn, Z., (2000). Situating vision in the world. *Trends in Cognitive Science, 4(5).* 197-207

Richardson, D.C, & Spivey, M.J., (2000) Representation, space and Hollywood Squares: Looking at things that aren’t there anymore. *Cognition, 76(3)* 269-295.

Richter, W., Somorjai, R., Summers, R., Jarmasz, M., Menon, R, Gati, J.S., Georgopoulos, A.P., Tegeler, C., Ugurbil, K., & Kim, S. (2000) Motor area activity during mental rotation studied by time-resolved single-trial fMRI. *Journal of Cognitive Neuroscience,*12(2), 310-320.

Shiffrar, M., & Freyd, J.J., (1993). Timing and apparent motion path choice with human body photographs.. *Psychological Science, 4(6),* 379-384.

Spivey, M.J, Tyler, M., Richardson, D.C. & Young, E., (2000). Eye Movements During Comprehension of Spoken Scene Descriptions. *Proceedings of the Twenty-second Annual Meeting of the Cognitive Science Society*, Erlbaum: Mawhah, NJ.

Stein, L.A., (1994). Imagination and situated cognition. Journal of Experimental and Theoretical Artificial Intelligence, 6.393-407.

Stevens, J.A., Fonlupt, P., Shiffrar, M., & Decety, J., (2000). New aspects of motion perception: Selective neural encoding of apparent human movements.. *Neuroreport: For Rapid Communication of Neuroscience Research, 11(1)*, 109-115.

Stoet & Hommel. (1999) Action planning and the temporal binding of response codes.. *Journal of Experimental Psychology: Human Perception and Performance*, 25(6), 1625-1640.

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

Wohlschläger & Wohlschläger. (1998) Mental and manual rotation. *Journal of Experimental Psychology: Human Perception and Performance, 2(2),* 397-412.