Available online at www.sciencedirect.com
ScienceDirect
Cognition 102 (2007) 476–485
COGNITION
www.elsevier.com/locate/COGNIT

Brief article

Representing object colour in language comprehension ☆

Louise Connell *

Cognition & Communication Research Centre, Division of Psychology, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK

Received 23 February 2006; accepted 27 February 2006

Abstract

Embodied theories of cognition hold that mentally representing something *red* engages the neural subsystems that respond to environmental perception of that colour. This paper examines whether implicit perceptual information on object colour is represented during sentence comprehension even though doing so does not necessarily facilitate task performance. After reading a sentence that implied a particular colour for a given object, participants were presented with a picture of the object that either matched or mismatched the implied colour. When asked if the pictured object was mentioned in the preceding sentence, people's responses were faster when the colours *mismatched* than when they *matched*, suggesting that object colour is represented differently to other object properties such as shape and orientation. A distinction between stable and unstable embodied representations is proposed to allow embodied theories to account for these findings.

Keywords: Embodied cognition; Language comprehension; Mental representation; Colour; Perception; Stability

☆ This manuscript was accepted under the editorship of Jacques Mehler.

* Tel.: + 44 191 227 3446; fax: +44 191 227 4515.
E-mail address: louise.connell@northumbria.ac.uk

0010-0277/$ - see front matter
doi:10.1016/j.cognition.2006.02.009

Embodied theories of cognition hold that thought is grounded in the same neural systems that govern sensation, perception and action (Barsalou, 1999; Glenberg & Kaschak, 2002; Johnson-Laird, 1983; Pecher & Zwaan, 2005). While there is currently no single view of embodied cognition, its theories share many characteristics and assumptions (see Wilson, 2002) and one of the most influential is Barsalou's (1999) Perceptual Symbol Systems. According to this theory, concepts are essentially partial recordings of the neural activation that arises during perceptual and motor experiences. These recordings (or "perceptual symbols") can later be re-enacted as a perceptual simulation of that concept. For example, if we think about having *soup* for lunch, the neural systems for vision, action, touch, taste, smell, etc. that were engaged during our previous experiences of *soup* are reactivated in a partial simulation. This simulation may include visual information of liquid in a bowl, sensorimotor information of eating hot savoury liquid with a spoon, and so forth. Barsalou thus argues that Perceptual Symbol Systems theory avoids transduction and grounding problems by assuming that conceptual representations are based on the same systems that are used for perception and action.

A growing body of empirical work has emerged in support of embodied representations of concepts. For example, it has been shown that "low-level" sensorimotor representations play a role in "high-level" cognitive processes, such as language comprehension and memory retrieval (Glenberg & Kaschak, 2002; Kaschak et al., 2005; Richardson, Spivey, Barsalou, & McRae, 2003; Solomon & Barsalou, 2001; Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002). Importantly, these studies employed implicit tasks, such as recognition and naming, demonstrating that perceptual information (e.g., object shape, orientation, motion) is activated even though doing so does not facilitate task performance. For example, Stanfield and Zwaan (2001) presented people with sentences that implied a particular object orientation (e.g., "John put the pencil in the drawer"), followed by a picture of the object (e.g., a pencil). People were faster to verify that a pencil had been mentioned when it was pictured in the orientation implied by the sentence (i.e., horizontally rather than vertically). The authors concluded that their findings could be explained by participants constructing an embodied representation of the sentence (e.g., a sensorimotor simulation of placing a pencil in a drawer), as this would include implied information about the pencil's orientation.

1. Representing colour information

Colour representation is a key aspect of perceptual information that has not received the same attention in the embodiment debate as other visual object attributes, such as shape, size and orientation (although see Chao & Martin, 1999). However, there is a long history behind the idea that object colour may be represented differently to other object properties. In the 17th century, the philosopher John Locke (1690/1975) argued for a distinction between *primary* and *secondary* object properties: primary properties (e.g., shape, size, motion,) were those that existed independent of us and could be perceived by multiple senses, while secondary properties (e.g., colour, taste, smell) were observer-dependent and could be perceived by only one of our senses (see also Jackson, 1977).

In more recent times, visual cognition research has provided some evidence to support the idea that colour, at least, differs from "primary" properties like shape, size and motion. For example, infants of 4.5 months individuate objects by their shape and size but cannot use colour to identify objects until the age of 11.5 months (Wilcox, 1999). In adults, Vandenbeld and Rensink (2003) found that memory for object colour is quite poor (rapid decay from 100 to 700 ms after exposure to a stimulus) compared to object shape (gradual decay after 100 ms) and size (gradual decay after 1900 ms). Electrophysiological evidence has also demonstrated the primacy of object shape over object colour (Proverbio, Burco, del Zotto, & Zani, 2004). When asked to respond to a specific colour such as *yellow*, Proverbio et al.'s participants found it difficult to ignore knowledge about object shape and showed greater brain response (selection negativity) when shown a canonically yellow object (e.g., *yellow chick*) than an arbitrarily yellow object (e.g., *yellow sweater*). However, when asked to respond to a specific shape such as *a piglet*, people found it easy to ignore knowledge about object colour and showed no response difference whether the object was pictured with canonical or arbitrary colour (e.g., same response to *pink piglet* as to *blue piglet*). The contrast between object colour and other object properties has also been highlighted by Aginsky and Tarr (2000), who showed that explicitly detecting changes in object colour required greater attentional resources than detecting changes in object position or presence. They concluded that colour is not as salient as other properties that determine the configuration of a scene (e.g., object presence, position, or shape) and hence that colour is encoded with less stability in scene representations.

2. The current study

This study aims to examine whether implicit perceptual information on object colour is represented during sentence comprehension. In the experiment, participants are presented with short sentences that imply (rather than explicitly state) a colour for a particular object. Each sentence is followed by a picture and participants are asked to indicate whether the pictured object was mentioned in the sentence. For test items, the pictured object was always mentioned in the preceding sentence but the object was shown in one of two picture conditions: *matching* the colour implied by the sentence or *mismatching* the colour implied by the sentence.

The embodied view holds that both sentence versions in Fig. 1 would be represented with the implied colour encoded as part of the perceptual simulation for *steak* – simulating a steak on a plate would involve specialising the steak colour to the appropriate *brown*, while simulating a steak in a butcher's window would involve specialising the steak colour to the appropriate *red*. However, there are two possible ways in which the match and mismatch conditions may differ in response latencies.

The first possibility assumes that object colour is represented in the same way as other object properties (Barsalou, 1999; Zwaan, 2004). This view would predict that

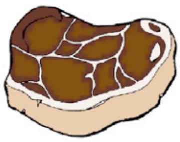
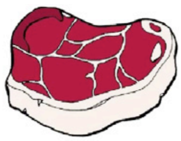
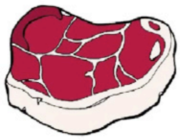
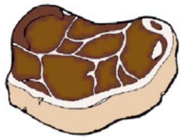
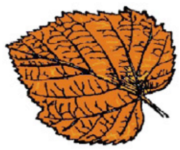
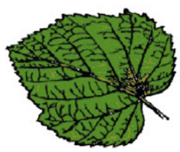
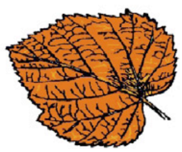
Sentence	Picture Condition: Match	Picture Condition: Mismatch
John looked at the steak on his plate.	 (brown steak)	 (red steak)
John looked at the steak in the butcher's window.	 (red steak)	 (brown steak)
Sarah stopped in the woods to pick a leaf off the ground.	 (orange leaf)	 (green leaf)
Sarah stopped in the woods to pick a leaf off a tree.	 (green leaf)	 (orange leaf)

Fig. 1. Sample sentence and picture stimuli used in experiment. Note that each colour used in both the match and mismatch picture conditions is a valid colour representation of that particular object (e.g., raw steaks are *red*, cooked steaks are *brown*). Full materials are available from the author on request. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.)

responses will be faster in the match condition because the representation constructed during sentence comprehension creates a pattern of activation in object colour processing areas that can facilitate the recognition of colour stimuli with a matching pattern of activation. Such *match facilitation* has been previously found for object shape (Zwaan et al., 2002), orientation (Stanfield & Zwaan, 2001) and motion (Zwaan, Madden, Yaxley, & Aveyard, 2004).

The second possibility assumes that object colour is represented differently to other object properties. Rather, since shape is more important than colour in object recognition (see Tanaka, Weiskopf, & Williams, 2001, for review), it is advantageous for participants in this task to ignore colour and attend to shape when trying to identify the pictured object. In the absence of any prior context, people can easily ignore colour when attending to shape in object recognition (Proverbio et al., 2004). However, it may prove more difficult to ignore the perceptual *red* of a pictured steak if a perceptual simulation of *redness* has been activated by the preceding sentence. Therefore, this view suggests slower responses in the match condition because, when a mismatching property is unstable, it is possible to ignore it in favour of stable shape with minimal cost. Such *mismatch facilitation* would represent a novel finding for an object property in this paradigm.[1]

There is another possibility – that no difference in response times will emerge between the match and mismatch conditions. Amodal theories of cognition assume that thought involves the representation and manipulation of discrete, amodal symbols (e.g., Kintsch & van Dijk, 1978; Fodor, 1975; Pylyshyn, 1984), and have been a cornerstone of symbolic cognitive modelling in a variety of domains (e.g., Anderson, 1993; Connell & Keane, 2006; Keane, Ledgeway, & Duff, 1994; Newell, 1990). In this view, people could simply be confirming that the pictured object (e.g., STEAK) was explicitly mentioned in the preceding sentence (e.g., LOOK(JOHN,STEAK[LOCATION:ON_PLATE])). Thus, a null effect would be more consistent with amodal views of representation than with embodied views.

2.1. Method

2.1.1. Materials

Forty-four pictures were created for use in this experiment. Of these, 24 were test items (forming pairs of pictures) and 20 were fillers (unrelated standalone pictures). Many of the pictures came from popular clipart packages but some were created by the author. All pictures were coloured naturalistically by sampling shades from photographs of the relevant objects, and contained only one predominant colour (e.g., Fig. 1's red steak predominantly contains shades of *red*). Each pair of test pictures was identical except for the colours used. All pictures were resized to a maximum of 250-pixel height (approximately 6.9 cm onscreen) and 350-pixel width (approximately 9.7 cm onscreen). Forty-four sentences were constructed to accompany the pictures. Of these, 24 were test items (naming an object featured in a test picture) and 20 were fillers (naming objects not featured in either test or filler pictures). The test sentences thus formed pairs, with each member of a pair implying a different colour for the same object. Filler sentences all contained at least one concrete noun (see Fig. 2).

Pictures were pretested in order to ensure that object recognition would not be affected by the canonicality or view specificity of the pictures. Each pair of test pictures (e.g., *red/brown* steak) was separated to form two groups of items

[1] It should be noted that Kaschak et al. (2005; see also Richardson et al., 2003) also found a mismatch facilitation for object motion (c.f. Zwaan et al., 2004). However, Kaschak et al. used a very different simultaneous presentation paradigm, and their explanation – that a visual stimulus "ties up" neural motion mechanisms and hinders ability to process a simultaneously presented auditory sentence describing motion in the same direction – cannot be used to predict effects in this study's sequential presentation paradigm.

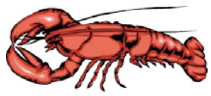
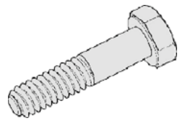
Filler Sentence	Picture	Comprehension Question
Brian picked up the rake that was lying in the garden.		The rake was in the garden?
Ben lost his briefcase somewhere on his way to work.		Ben found his briefcase?

Fig. 2. Filler sentence, picture and comprehension question stimuli. All filler pictures were unrelated to the preceding sentence. Comprehension questions followed half of filler trials and required an even distribution of yes/no answers.

that were presented onscreen after their object names, while filler pictures appeared in both groups and were presented after semantically unrelated words. Twenty-two participants, randomly assigned to each group, were asked to indicate whether each picture matched the preceding name and then rate the picture on a scale from 1 (poor quality) to 7 (good quality). All pictures used as test items in this experiment had a median response time <1250 ms ($M = 866$, $SD = 195$) with no significant response-time difference between the pictures in each test pair ($ps > 0.2$). In addition, all pictures received a median quality rating of at least 4/7 (test $M = 4.58$, $SD = 0.54$; filler $M = 4.68$, $SD = 0.50$). Test sentences were also pretested to ensure they actually implied the intended colour for the object. Each pair of test sentences was separated to form two groups of items that were presented along with both matching and mismatching pictures. Twenty-four new participants were given four forced-choice alternatives stating that each sentence was best matched by: (a) the first picture, (b) the second picture, (c) both pictures equally, or (d) neither picture. All test sentences had the picture from the matching condition chosen at least 50% of the time ($M = 84\%$, $SD = 13\%$).

2.1.2. Design

Test items were divided into four groups so that each group featured one of four sentence–picture combinations (see Fig. 1). Each group contained equal numbers of match and mismatch test items, and the various colours featured in test pictures were distributed approximately evenly across groups. Participants were assigned randomly to one of the groups. Thus, the experiment was a 2 (sentence version: versionl, version2) × 2 (picture condition: match, mismatch) × 4 (group) design, with sentence version and picture condition as within-participants variables and group as a between-participants variable.

2.1.3. Participants

Forty-four native speakers of English from Northumbria University (not involved in pretests) were paid a nominal sum for participation in this experiment.

Table 1
Mean response times (ms) and proportion of errors, with standard deviations in parentheses

Picture condition	Response times	Errors
Match	1369 (638)	0.074 (0.262)
Mismatch	1215 (509)	0.071 (0.258)

2.1.4. Procedure

Participants read instructions describing the experiment that asked them to read every sentence carefully as their comprehension would be tested at various points during the experiment, and to respond as quickly as possible as their response time was being measured. Testing took place on portable computers running Presentation software. Each trial began with a left-aligned vertically centred fixation cross for 1000 ms, followed by presentation of a sentence. After pressing the spacebar to indicate comprehension, another fixation cross was displayed centrally onscreen for 500 ms, followed by a picture. Participants were asked to decide if the pictured object had appeared in the preceding sentence and indicate their decision by pressing the key labelled "yes" (comma key) or the key labelled "no" (full-stop key). Half of all filler trials had a comprehension question (relating to the filler sentence) appear after the picture, with participants answering equal numbers of "yes" and "no" questions (see Fig. 2).

2.2. Results and discussion

Table 1 shows the mean correct response times and accuracy for the match and mismatch picture conditions. Two participants who responded incorrectly to >25% of picture items were eliminated from the analysis. All responses <300 ms and >3000 ms were considered outliers and excluded, as were any responses more than two standard deviations outside a participant's mean in the relevant condition. Altogether, 9.5% of the data was excluded. Analyses of variance were run on the remaining data by participants and by items.

Results were consistent with the embodied rather than amodal view of representation, and consistent with the view that object colour is represented differently to other object properties. Against the assumption that object colour is represented in the same way as other object properties, people responded more quickly when the picture colour *mismatched* the object colour implied by the sentence than when it *matched*,[2] $F_1(1,38) = 9.458$, $MSE = 95887$, $p < 0.005$; $F_2(1,32) = 5.279$, $MSE = 57482$, $p < 0.05$. Response times showed no significant effect of sentence version [$F_1(1,38) = 2.485$, $MSE = 105097$, $p > 0.1$; $F_2 < 1$], and no interaction between sentence version and picture condition [$F_1 < 1$; $F_2 < 1$]. There was no difference in the proportion of errors between match and mismatch picture conditions [$F_1 < 1$; $F_2 < 2$], nor between sentence versions [$F_1(1,38) = 2.220$, $MSE = 0.029$, $p > 0.1$;

[2] This match > mismatch effect was also found in a previous version of this experiment (see Connell, 2005).

$F_2 < 1$], nor was there any significant interaction [$F_1(1,38) = 2.381$, $MSE = 0.036$, $p > 0.1$; $F_2(1,32) = 1.025$, $MSE = 0.017$, $p > 0.3$].

All effects involving the group variable were non-significant (ps > 0.1), with two exceptions where interactions in the proportion of errors were significant by participants only. The first was group × sentence [$F_1(3,38) = 3.043$, $MSE = 0.029$, $p < 0.05$; $F_2(3,32) = 1.576$, $MSE = 0.017$, $p > 0.2$] and the second was group × sentence × match [$F_1(3,38) = 2.957$, $MSE = 0.036$, $p < 0.05$; $F_2(3,32) = 1.213$, $MSE = 0.017$, $p > 0.3$], both due to higher errors for sentence-versionl in the match condition and for version2 in the mismatch condition within a single group. Due to lack of theoretical significance, these group effects will not be discussed further.

3. General discussion

The findings reported in this paper are consistent with the idea that language comprehension involves constructing sensorimotor simulations of a described scenario, where such an embodied representation includes information not explicitly stated. Results showed that perceptual colour information is activated during sentence comprehension even though doing so does not facilitate task performance: people responded more quickly when the colour of a pictured object *mismatched* the colour implied by the previous sentence. This finding is of particular interest as it runs contrary to that predicted by current embodied theories which hold that matching implied information should facilitate faster responses[1]. This paper therefore proposes a distinction between stable and unstable embodied representations as explanation for the results.

Work in visual cognition has found that colour is not as salient as other properties that determine the configuration of a scene and hence is encoded with less stability in scene representations (Aginsky & Tarr, 2000). In terms of Barsalou's Perceptual Symbol Systems, these results suggest that there may be differences in how various properties are specialised in a perceptual simulation of a concept. Many of Locke's (1690/1975) primary properties – such as shape, size, motion – are multimodal (can be perceived by multiple senses) and highly salient in configuring a visual field: for this reason, it may make evolutionary sense that visually oriented humans represent them as *stable* specialisations of perceptual simulations. In turn, many of Locke's secondary properties – colour, taste, smell – are unimodal (can be perceived by only one sense) and are not salient in visual field configuration: again, it may make evolutionary sense that people might represent them with less stability and ignore them when necessary.

Since shape is more important in object recognition than either colour (Tanaka et al., 2001) or orientation (e.g., Harris & Dux, 2005), it is useful to contrast the present study (examining one unstable and one stable property: colour and shape) with another study using the same methodology (Stanfield & Zwaan, 2001, examining two stable properties: orientation and shape). When perceptual input mismatches perceptual simulation on an *unstable* property (e.g., seeing a red steak after reading about a brown steak), there is minimal interference as people can easily ignore this unimportant, unimodal mismatch. Accordingly, when perceptual input matches perceptual simulation, the matching colour information is somewhat difficult to ignore because neural subsystems simulating this colour are already active in the perceptual simulation, and so it interferes more with shape recognition. This reflects the finding of the present paper that the match condition was slower than the mismatch condition. In contrast, when perceptual input mismatches perceptual simulation on a *stable* property (e.g., seeing a vertical pencil after reading about a horizontal pencil), then there is a processing cost as the participant tries to ignore the jarring presence of a multimodal mismatch. This reflects Stanfield and Zwaan's (2001) finding that the mismatch condition was slower than the match condition.

While this account is somewhat speculative, it represents one possible explanation of an interesting and counterintuitive effect. Further research will be needed to test the impacts of the proposed distinction between stable and unstable object specialisations, but it is hoped that such a distinction will provide a novel and useful perspective in understanding how embodied representations are formed.

Acknowledgements

This work was funded by the Division of Psychology, Northumbria University. Many thanks to Dermot Lynott for valuable discussion, to Ben Singleton and Darren Dunning for data collection, and to anonymous reviewers for comments on a previous version of this study.

References

Aginsky, V., & Tarr, M. J. (2000). How are different properties of a scene encoded in visual memory? *Visual Cognition, 7*, 147–162.

Anderson, J. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral & Brain Sciences, 22*, 577–660.

Chao, L. L., & Martin, A. (1999). Cortical regions associated with perceiving, naming, and knowing about colors. *Journal of Cognitive Neuroscience, 11*, 25–35.

Connell, L. (2005). Color and stability in embodied representations. In B. Bara & L. Barsalou (Eds.), *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society* (pp. 482–487). Hillsdale, NJ: Erlbaum.

Connell, L., & Keane, M. T. (2006). A model of plausibility. *Cognitive Science, 30*(1), 95–120.

Fodor, J. (1975). *The language of thought*. Cambridge: Harvard University Press.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*, 558–565.

Harris, I. M., & Dux, P. E. (2005). Orientation-invariant object recognition: evidence from repetition blindness. *Cognition, 95*(1), 73–93.

Jackson, F. (1977). *Perception: A representative theory*. Cambridge: Cambridge University Press.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

Kaschak, M. P., Madden, C. J., Thernault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., & Zwaan, R. A. (2005). Perception of motion affects language processing. *Cognition, 94*, B79–B89.

Keane, M. T., Ledgeway, T., & Duff, S. (1994). Constraints on analogical mapping: a comparison of three models. *Cognitive Science, 18*, 387–438.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363–394.
Locke, J. (1975). An essay concerning human understanding. In P.H. Nidditch (Ed.), Oxford: Clarendon Press. (Original work published 1690).
Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
Pecher, D., & Zwaan, R. A. (2005). Introduction to grounding cognition. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: the role of perception and action in memory, language, and thinking*. Cambridge: Cambridge University Press.
Proverbio, A. M., Burco, F., del Zotto, M., & Zani, A. (2004). Blue piglets? Electrophysiological evidence for the primacy of shape over color in object recognition. *Cognitive Brain Research, 18*, 288–300.
Pylyshyn, Z. W. (1984). *Computation and cognition*. Cambridge, MA: MIT Press.
Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science, 27*, 767–780.
Solomon, K. O., & Barsalou, L. W. (2001). Representing properties locally. *Cognitive Psychology, 43*, 129–169.
Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12*, 153–156.
Tanaka, J., Weiskopf, D., & Williams, P. (2001). The role of color in high-level vision. *Trends in Cognitive Sciences, 5*, 211–215.
Vandenbeld, L. A., & Rensink, R. A. (2003). The decay characteristics of size, color, and shape information in visual short-term memory. *Journal of Vision, 3*, 682.
Wilcox, T. (1999). Object individuation: Infants' use of shape, size, pattern, and color. *Cognition, 72*, 125–166.
Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*, 625–636.
Zwaan, R. A. (2004). The immersed experiencer: toward an embodied theory of language comprehension. In B. H. Ross (Ed.). *The psychology of learning and motivation* (Vol. 44, pp. 35–62). New York: Academic Press.
Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: dynamic representations in language comprehension. *Cognition, 28*, 611–619.
Zwaan, R. A., Stanfield, R. A., & Yaxley, RH (2002). Language comprehenders routinely represent the shapes of objects?. *Psychological Science, 13*, 168–171.