Individual differences in spatial cognition influence mental simulation of language

Nikola Vukovic [a,b,*], John N. Williams [a]

[a] Department of Theoretical and Applied Linguistics, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DA, United Kingdom
[b] Centre for Cognition and Decision Making, National Research University Higher School of Economics, Moscow, Russia

A B S T R A C T

The factors that contribute to perceptual simulation during sentence comprehension remain underexplored. Extant research on perspective taking in language has largely focused on linguistic constraints, such as the role of pronouns in guiding perspective adoption. In the present study, we identify preferential usage of egocentric and allocentric reference frames in individuals, and test the two groups on a standard sentence-picture verification task. Across three experiments, we show that individual biases in spatial reference frame adoption observed in non-linguistic tasks influence visual simulation of perspective in language. Our findings suggest that typically reported grand-averaged effects may obscure important between-subject differences, and support proposals arguing for representational pluralism, where perceptual information is integrated dynamically and in a way that is sensitive to contextual and especially individual constraints.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The mind sciences have recently witnessed a dramatic shift in the way we conceptualise higher mental functions – of which language is a paradigmatic example – wherein these are fundamentally linked to more basic cognitive and neurobiological mechanisms. According to this embodied account, language comprehension involves simulation of content analogous to that arising during our perceptual, motor, and emotional experiences (Barsalou, 2008, 2010). For example, a phrase such as "writing a letter" is interpreted by the brain through a partial recreation of what it would be like for us to perform that action. At the same time, theories of action understanding are pointing out that mechanisms used in interpreting others' actions share a common representational space with mechanisms used during real action execution (Avanzini et al., 2012; Avenanti, Candidi, & Urgesi, 2013; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). However, in context, interpreting language such as the phrase above can be achieved through the use of different perspectives or reference frames. On the one hand, one could adopt an egocentric, or internal, perspective and imagine performing the action oneself. Equally, one could assume the role of an external observer of the action – in other words, an allocentric perspective. It is clear that the use of egocentric and allocentric strategies leads to the construction of completely different representations of space and acting entities within it, therefore a successful theory must specify how perspective taking is achieved in understanding. Which factors determine perspective adoption during the simulation of comprehended language remains, however, a surprisingly underexplored topic.

The most straightforward results in relation to perspective come from studies of single word reading. There, the consensus appears to be that comprehenders adopt a first person, egocentric, perspective. In studies where participants were reading action verbs such as *run*, *punch*, or *talk*, activation has been seen in areas of the motor cortex which selectively control leg, arm, and face movements (Pulvermüller, 2011; Pulvermüller, Shtyrov, & Ilmoniemi, 2005). That is, interpreting action words was in these participants neurophysiologically similar to actually performing the respective movements themselves. Further MEG and EEG studies have shown that adopting the role of an embodied first-person agent is a very rapid process, with motor activation following single action word reading occurring within 200 ms of stimulus onset (Boulenger, Shtyrov, & Pulvermüller, 2012; Shtyrov, Butorina, Nikolaeva, & Stroganova, 2014). Additional evidence that word comprehension proceeds through a simulation from the perspective of an embodied actor, rather than a patient or observer, comes from experiments investigating the notion of body-specificity.

* Corresponding author at: Department of Theoretical and Applied Linguistics, English Faculty Building, University of Cambridge, 9 West Road, Cambridge CB3 9DP, United Kingdom.

E-mail address: nv254@cam.ac.uk (N. Vukovic).

Willems, Hagoort, and Casasanto (2010) used fMRI scanning to test left-handers and right-handers in a lexical decision task. They found that manual-action verbs, compared to nonmanual ones, caused increased activity in contralateral premotor brain areas. What this meant was that left-handers preferentially activated the right motor cortex, whereas right-handers preferentially activated corresponding areas in the left hemisphere. To confirm that this activation was functionally relevant, Willems, Labruna, D'Esposito, Ivry, and Casasanto (2011) did a follow-up study using Transcranial Magnetic Stimulation (TMS) where they found that right-handed participants were faster to perform a lexical decision on hand-related action verbs (but not other action verbs) when stimulation was applied to the left (rather than the right) premotor cortex. All of these results provide support for the idea that people understand action words by simulating performing the actions themselves, in a manner which is consistent with their specific body characteristics (such as, for example, handedness).

As informative as the above studies are, they provide only a glimpse into the complex dynamics of language processing, which we cannot fully understand by relying on the analysis of decontextualised words. As mentioned, these studies support the notion of an egocentric perspective bias during isolated word comprehension, however, our daily language usage is much more richly situated. One question is whether people continue to interpret language through the perspective of a first-person embodied agent even when reading sentences which can refer to multiple people. The evidence here is mixed, with some studies showing an egocentric preference, and others arguing for a more flexible interpretation. In the study of Tettamanti et al. (2005), Italian subjects listened to action sentences such as "I kick the ball" or "I bite an apple", and abstract control sentences such as "I appreciate sincerity". The findings revealed that action sentences, unlike abstract ones, produced stronger activity in fronto-parietal motor areas, where the relevant arm/leg/face movements are motorically coded. This study, and similar studies such as the one by Boulenger, Hauk, and Pulvermüller (2009), supports the idea previously expressed in research on single word processing, stating that people interpret action language from the perspective of an embodied agent. It has recently been pointed out, however, that many of the studies reported in the literature do not explicitly test perspective taking *per se*, and often conflate the perspective of the agent and the perspective of the thematic role assigned to a potentially self-referential pronoun, such as "I" (Beveridge & Pickering, 2013). A more direct contrast between the two has been drawn in studies which manipulate linguistic factors such as the number of thematic roles and personal pronouns used. Glenberg and Kaschak (2002) devised a method (ACE: action-sentence compatibility effect) where they asked participants to judge the sensibility of a sentence by performing a movement away from or towards their body. They observed that participants found it easier to judge the sentence as sensible when the action described matched the direction of the response. For example, having read the sentence "You delivered the pizza to Andy" participants were faster to respond by moving the hand away from their body. Conversely, they were slower to respond with the same movement after reading "Andy delivered the pizza to you". This study suggests that an egocentric perspective was adopted when the pronoun "you" also occupied the role of the agent in the sentence – otherwise, the simulation was performed from the thematic role assigned to the pronoun, namely the patient.

A study of Brunyé, Ditman, Mahoney, Augustyn, and Taylor (2009) investigated the role that first, second, and third person pronouns have on the mental simulation of meaning. The authors presented participants with short action sentences, such as "You are/She is cutting a tomato", following which they had to verify an image. Crucially, the image would either match an egocentric perspective (the position of the actor's hands matched the orientation of the participant), or it would present an external scene (as if looking at someone else's hands performing the action). They found that there is variability in the adoption of perspective, with the second person pronoun ("you") leading to egocentric perspective, and third person pronouns leading to an allocentric/external perspective. Interestingly, they were able to modulate the perspective adopted with the first person pronoun "I" to either internal or external, by presenting a prior descriptive narrative about the agent. Gianelli, Farnè, Salemme, Jeannerod, and Roy (2011) also found that second (but not third) person perspectives produced an ACE effect, with participants being faster to perform movements compatible with sentence implied movements. This pattern of data is further replicated in the findings of Sato and Bergen (2013), who confirmed that pronoun choice influences the simulated perspective: sentences about "you" facilitate verification of internal images, whereas sentences using third person pronouns lead to faster responses to images from an allocentric/external perspective. Interestingly, Ruby and Decety (2001) found that the left inferior parietal and somatosensory cortex distinguishes the egocentric and allocentric/external perspectives at the neural level. In their study, the presentation of sentences such as "I am stapling a sheet of paper" led participants to imagine the action from an external perspective (that of the experimenter). In comparison, using the pronoun "you", led to an internal simulation which was characterised by increased activity in somatosensory areas, in accordance with embodied accounts.

To sum up, the above studies suggest differential perspective adoption as a function of pronoun choice in sentences. However, it is clear that there is much variability with respect to when egocentric and allocentric perspectives are adopted. Clearly, more data is needed in order to draw definitive conclusions about the factors driving the choice of reference frames during comprehension.

At the same time as embodied cognition research on word meaning has urged for consideration of processes external to the abstract linguistic system, research done on sentence-level comprehension has focused almost exclusively on linguistic factors, such as the role of pronouns on the selection of spatial reference frames. This is surprising given that one of the key questions for embodied cognition is how our behaviour and thinking is influenced by the constraints of our own individual bodies and environmental contexts. The work reviewed above has given no consideration to individual differences in perspective-taking strategies and preferences, despite the fact that such differences have been demonstrated in numerous experimental tasks (for an overview, see Gramann, 2011). For example, different people have been shown to preferentially use egocentric or allocentric reference frames during navigation, environmental learning, as well as other spatial and nonspatial tasks (Lawton, 1994, 1996; Taylor & Tversky, 1996). We argue that a consideration of differences in reference-frame selection is crucial when investigating embodied mental simulation of action, particularly having in mind the role spatial reference frames have in language comprehension. The egocentric frame places the agent at the centre of the coordinate system, with spatial and entity information processed based on the agent's orientation. The allocentric frame, on the other hand, organises and describes relations between objects externally, independent of the agent's perspective (Klatzky, 1998). Recently, Beveridge and Pickering (2013) have laid out theoretical considerations for why a selection of spatial reference frames (be they egocentric or allocentric) is indeed a prerequisite for embodied action simulation. According to this view, a comprehender can only perform an embodied simulation of action when it is grounded in a spatial context, which in turn is provided by the comprehender's situation model or frame of reference. For example, if when reading the sentence "I am kicking you" a person selects an egocentric