Cognition 147 (2016) 75–84

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Listeners use speaker identity to access representations of spatial perspective during online language comprehension

Rachel A. Ryskin [a,*], Ranxiao Frances Wang [a,b], Sarah Brown-Schmidt [a,b]

[a] Department of Psychology, University of Illinois at Urbana-Champaign, United States
[b] Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, United States

ARTICLE INFO

Article history:
Received 8 November 2014
Revised 5 November 2015
Accepted 20 November 2015
Available online 27 November 2015

Keywords:
Spatial perspective-taking
Partner-specific encoding
Language comprehension
Eye-tracking

ABSTRACT

Little is known about how listeners represent another person's spatial perspective during language processing (e.g., two people looking at a map from different angles). Can listeners use contextual cues such as speaker identity to access a representation of the interlocutor's spatial perspective? In two eye-tracking experiments, participants received auditory instructions to move objects around a screen from two randomly alternating spatial perspectives (45° vs. 315° or 135° vs. 225° rotations from the participant's viewpoint). Instructions were spoken either by one voice, where the speaker's perspective switched at random, or by two voices, where each speaker maintained one perspective. Analysis of participant eye-gaze showed that interpretation of the instructions improved when each viewpoint was associated with a different voice. These findings demonstrate that listeners can learn mappings between individual talkers and viewpoints, and use these mappings to guide online language processing.

1. Introduction

Much of human communication requires keeping track of what another person knows. For example, when a coworker says, "How was the talk?", taking the sentence at face-value you might begin to think about every talk that you have ever attended (or even heard of), which would lead you to an uninformative response—at best a clarification question, at worst telling your coworker about something irrelevant. In the case of something that occurs more frequently than "talks", it could even lead to an interminable memory search process. However, a more effective strategy, and one that successful communicators must employ, takes into account what your coworker knows and narrows the search space to only talks that she knows occurred, that she knows she did not attend, and that she knows you did attend. Tracking others' knowledge places constraints on the set of possible intended referents and eases the burden of comprehension, allowing conversation to proceed smoothly. The nature of this constraining knowledge can take many forms, from what topics have been previously discussed between two interlocutors, to an individual's viewpoint on the physical environment.

In order to understand a speaker, listeners must consider that speaker's perspective (Clark, 1992) and how it may differ from their own. Indeed, listeners are sensitive to differences in perspectives between themselves and an interlocutor and bring this information to bear in the early moments of processing a sentence (Brown-Schmidt, 2009, 2012; Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Hanna, Tanenhaus, & Trueswell, 2003; Heller, Grodner, & Tanenhaus, 2008; Nadig & Sedivy, 2002). The bulk of this evidence comes from paradigms in which a difference in perspectives between the speaker and listener is created by occluding an item from the speaker's view. Much of this research shows that listeners (at least partially) discount occluded objects as potential referents, on the assumption that the speaker is unlikely to speak about something they have no knowledge of. This successful use of perspective corresponds to what has been referred to as Level 1 knowledge—mental simulation that involves distinguishing what is visible to oneself from what is visible to others, as in occlusion situations. Level 1 knowledge emerges early in development and is thought to require little cognitive effort, even by age three (Flavell, Everett, Croft, & Flavell, 1981; Masangkay et al., 1974).

Differences in perspective can arise from situations other than occlusion, as well. In particular, differing spatial viewpoints, which are the focus of the present manuscript, require interlocutors to take this into account in order to understand each other (Schober, 1993). It has been argued that this Level 2 knowledge—the ability to appreciate not only that another person sees something, but how they see it—emerges later in development and is more cognitively effortful (Apperly & Butterfill, 2009; Flavell et al., 1981; Salatas & Flavell, 1976).

* Corresponding author at: Department of Psychology, University of Illinois at Urbana-Champaign, 603 E. Daniel St., Champaign, IL 61820, United States.
E-mail address: ryskin2@illinois.edu (R.A. Ryskin).

http://dx.doi.org/10.1016/j.cognition.2015.11.011
0010-0277/

In daily life, the spatial viewpoints of conversation partners are often misaligned. In fact, in some sense, they are never truly aligned because interlocutors can never inhabit the exact same location at the same time (Schober, 2009). Speakers frequently and spontaneously take into account such differences in perspectives when communicating (Tversky & Hard, 2009). For instance, when giving walking directions to a friend, you might say, "From Sixth Street you'll take a *left* on Daniel Street, and I'll be standing halfway up the block." From your own perspective, your location is actually to the *right* of Sixth Street but, as a courtesy to your friend who is unfamiliar with the area, you take their perspective in order to avoid confusion. Indeed, Schober (2009) found that when participants with high spatial perspective-taking ability are matched with participants with low abilities, they adopt spatial language consistent with that partner's perspective more often when giving directions compared to when they are paired with someone who is equally capable at spatial perspective-taking. Similarly, speakers use their egocentric perspective less when directing a person who is unable to provide immediate feedback about whether they understood the spatial instruction (Schober, 1993). The fact that speakers often choose to start with their own perspective when they know that any confusion can be easily resolved (i.e., when the person receiving the instructions can ask for clarification) points to the inherent difficulties of performing a spatial perspective transformation.

In this paper, we briefly review what is known about how spatial perspectives are represented in memory and the mental computations required to imagine another perspective. We then discuss how these spatial memory representations might be called upon during language processing when speaker and listener perspectives differ. We hypothesize that spatial perspectives of interlocutors can be tied to the speaker's identity in memory and accessed on-line to constrain interpretation of what has been said. Finally, we provide empirical evidence that supports our hypothesis and discuss the implications of our findings for theories of perspective-taking and language comprehension more broadly.

1.1. Spatial perspective-taking and memory

Studies of memory for spatial layouts of objects indicate that a mental change in viewpoint renders information about object-to-object relations more difficult to retrieve compared to when the viewpoint remains stable (e.g., Shelton & McNamara, 1997, 2001; Simons & Wang, 1998). Rieser (1989) asked participants to memorize an array of objects and then tested their ability to retrieve the relative spatial location of an object from a novel point of view. Participants had more difficulty doing so when the novel location was reached by a rotation than by a simple translation. One potential explanation for this processing cost associated with viewpoint rotation comes from evidence that participants most often encode the environment, and objects within it, using an egocentric reference frame (e.g., Wang, 2007, 2012). As a result, taking another perspective requires the effortful transformation of the original (egocentric) reference frame to fit a new orientation (Easton & Sholl, 1995; Kessler & Thomson, 2010; Mou, McNamara, Valiquette, & Rump, 2004). Others have argued that the processing cost results primarily from the sensorimotor interference created between the coordinates in the person's own perspective and those in the imagined perspective (Brockmole & Wang, 2003; May, 2004; Wang, 2005).

Furthermore, the difficulty of spatial perspective-taking increases with the angular disparity between the participant's viewpoint and the novel viewpoint presented at test (e.g., Huttenlocher & Presson, 1973; Kessler & Rutherford, 2010; Kessler & Thomson, 2010; Levine, Jankovic, & Palij, 1982; Rieser, 1989; Surtees, Apperly, & Samson, 2013). The detrimental effects of greater angular disparity suggest that spatial perspective-taking is an embodied cognitive process (Kessler & Thomson, 2010). Thus, a listener taking into account the perspective of her conversation partner will mentally rotate her egocentric perspective to align it with the partner's.

One way to reduce the cognitive burden of spatial perspective-taking is by providing advance information about a viewpoint. Studies asking participants to imagine a perspective before seeing an array from the new viewpoint show that the representation of a perspective can be maintained in memory in the absence of the visual array that it applies to (Avraamides, Ioannidou, & Kyranidou, 2007; Avraamides, Theodorou, Agathokleous, & Nicolaou, 2013; c.f. Wang, 2005). Further, Galati, Michael, Mello, Greenauer, and Avraamides (2013) provide evidence that speakers do learn and store representations of their future conversation partner's spatial viewpoint, when it is made available to them in advance. However, the nature of these representations and how they are stored and accessed may differ substantially between speakers and listeners. The task of the speakers is to put into words the spatial perspective that they have chosen to adopt whereas the listeners must remain flexible enough in their representations to adapt to whichever unknown perspective they are about to hear an instruction from.

1.2. Spatial perspective-taking during language processing

The challenges involved in representing others' spatial perspectives are well documented. Yet, the comprehension processes involved in interpreting spatial language from a perspective that differs from one's own are less well understood. It is clear that speakers often produce spatial language from the intended recipient's perspective and listeners (or readers) can come to understand spatial directions that are given from a different perspective (Schober, 1993; Tversky & Hard, 2009; Taylor & Tversky, 1992). Yet, little is know about the mechanisms involved in, or the time-course of, adopting a different spatial perspective during language comprehension.

The integration of an occlusion-based difference in perspectives occurs rapidly during sentence interpretation (e.g., Heller et al., 2008). However, the processes involved in computing a differing perspective are not the same when that difference is the result of occlusion compared to when it stems from an alternative spatial orientation (Michelon & Zacks, 2006). Occlusion prompts participants to use a simple line-of-sight tracing strategy to compute the differences between their perspective and that of their partner. By contrast, when spatial perspectives are misaligned, participants must undergo an imagined transformation of their perspective and remapping of reference frames, which may lead to a conflict between the imagined and egocentric reference frames. Thus, spatial perspective-taking and occlusion-based perspective-taking may differently guide the on-line comprehension of utterances.

Nonetheless, some evidence suggests that listeners are able to use information about a speaker's spatial viewpoint to constrain the interpretation of a sentence as it unfolds. Ryskin, Brown-Schmidt, Canseco-Gonzalez, Yiu, and Nguyen (2014) monitored the eye movements of listeners as they processed sentences with potentially ambiguous spatial language. Participants heard instructions to move objects around a complex display of animals with accessories (e.g., a hat, a purse). The instructions, such as "*Go left to* <u>*the pig with the hat*</u>," were given either from the participant's egocentric perspective (i.e., "left" = participant's left) or the opposite perspective (a 180° rotation; "left" = participant's right). The displays were designed such that instructions were temporarily ambiguous between two potential referents. For example, "*the pig with the*..." was temporarily consistent with two different pigs