A Cognitive Model for Spatial Perspective Taking

Laura M. Hiatt (lahiatt@stanfordalumni.org)
CSLI, Stanford University
220 Panama Street, Stanford, CA 94305 USA

J. Gregory Trafton (trafton@itd.nrl.navy.mil)
Naval Research Laboratory
4555 Overlook Avenue, Washington, DC 20375 USA

Anthony Harrison (anh23@pitt.edu)
Learning Research and Development Center, University of Pittsburgh
3939 O'Hara St., Pittsburgh, PA 15260 USA

Alan C. Schultz (schultz@aic.nrl.navy.mil)
Naval Research Laboratory
4555 Overlook Avenue, Washington, DC 20375 USA

Introduction

When communicating with other people, one of the basic things that people must do is take others' perspectives. Most of the experimental work on spatial language and perspective taking has focused on four frames of reference: exocentric (world-based, such as "Go north"), egocentric (self-based, "Turn to my left"), addressee-centered (other-based, "Turn to your left") and object-centric (object-based, "The fork is to the left of the plate") (Carson-Radvansky & Logan, 1997; Carson-Radvansky & Radvansky, 1996; Levelt, 1984). Any time egocentric or addressee-centered frames of references are used, spatial perspective taking is needed: egocentric utterances require the listener to take the speaker's perspective, and addressee-centered utterances require the speaker to have already taken the listener's perspective. As part of a project to make intelligent agents and robots more useful to people, we have been developing cognitive models of spatial cognition and perspective taking (Trafton, Schultz, Cassimatis et al., under review; Trafton, Schultz, Perzanowski et al., under review).

Unfortunately, there are relatively few computational cognitive models of spatial cognition available in order to implement spatial perspective-taking models. One recent entrée to spatial cognition research has been ACT-R/S (Harrison & Schunn, 2003).

ACT-R/S extends ACT-R (Anderson & Lebiere, 1998) to implement a theory about spatial reasoning. It posits that spatial representations of objects are egocentric and dynamically updated (Wang & Spelke, 2002). ACT-R/S represents objects using vectors to the visible sides of the object. It has the ability to track these objects through a configural buffer, a data structure analogous to the other buffers of ACT-R that stores each object once it has been identified. The coordinate vectors of the objects in the buffer are then dynamically updated as the agent moves throughout the spatial domain. The configural buffer, unlike the visual and retrieval buffers of ACT-R, can hold more than one object to account for the fact that animals have been shown to track more than one landmark at once while moving through the world. The other spatial buffers within ACT-R/S, the visual and manipulative buffers, are not a central part of the work described here and are described more fully elsewhere (Harrison & Schunn, 2003).

Perspective Taking Model

In order to demonstrate the results of perspective taking using ACT-R/S, a simple 'fetch' task was designed for the model to perform in a simulated world. In this world, two agents (hereafter referred to as the 'speaker' and the 'robot') are in a room with two wrenches and a screen. In many cases, an utterance is ambiguous given the listener's knowledge, but unambiguous given the speaker's knowledge. Figure 1 is an example. The figure shows a robot and a person facing each other. The robot can see that there are two wrenches in the room, wrench1 and wrench2, but the person only knows about wrench2 because wrench1 is hidden from her. If a person said, "Robot, give me the wrench" (which is understood by our robotic system), the phrase "the wrench" is potentially ambiguous to the robot because there are two wrenches, though unambiguous to the person because she only knows of the existence of one wrench. Intuitively, if the robot could take the perspective of the person in this task, it would see that, from that perspective, wrench2 is the only wrench and therefore "the wrench" must refer to wrench2. Our model of spatial perspective taking uses ACT-R/S to accomplish this task. There are several components to perspective taking that the model goes through in order to successfully accomplish its goals.

Perspective taking process

The production rules involved in the perspective-taking process are the most important part of the model, as they implement the heart of its theory of spatial perspective taking. Taking the perspective of someone at position and orientation B, from position and orientation A, the over all procedure is to: (1) Turn to face position B; (2) Walk to position B; (3) Face orientation B; (4) Extract the desired information from the visual knowledge at this position and orientation; (5) Face position A; (6) Walk back to position A; and (7) Return to orientation A.

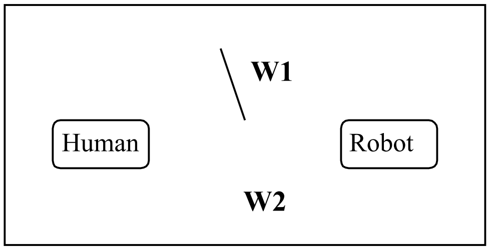

Figure 1: The human can only see one wrench (shown as W2), while the robot can see two.

The key to this process is that all of these movements – i.e. turning and walking – are *mentally* done by only transforming the configural buffer contents by the appropriate vector, leaving everything else the same. Thus the physical location of the robot does not change; it is only its mental location and perspective that changes.

Initial scan for objects The model first uses perspective taking to deduce where it should begin looking for the wrench. If the speaker gives a references such as "in front of me", or "to my left", the robot can use that information to constrain its search. It takes the speaker's perspective and then looks in front of it, or to its left (as indicated by the speaker's initial instructions), and keeps track of that location as it returns to its own perspective. This is where it begins its search for the wrench.

Deciding which wrench to give to the speaker The model also uses perspective taking once a wrench has been found. When it has located a wrench in the desired location, it looks around for obstacles that could possibly block the speaker's view of the wrench. If it finds any such obstacles, it takes the speaker's perspective again in order to judge whether or not the speaker can see that particular wrench from her perspective.

This time, however, once the robot has taken the speaker's perspective, instead of turning to match the speaker's orientation, it turns to face the located wrench. Determining whether or not the wrench is visible by the speaker is then done by comparing the transformed location vectors of the target object with the location vectors of the possible obstacles, making sure that the obstacle's vectors do not completely surround the target object's vectors. This ensures that the speaker has the ability to see at least part of the wrench.

If the speaker can in fact see the wrench, the robot hands that wrench to the speaker. If the speaker cannot, the robot continues to look for a wrench that the speaker can see.

Discussion

Using a person's (hypothesized) representation allows the cognitive agent to undergo perspective taking by imagining movement throughout the world by simply altering the representation of the objects in the configural buffer. This ultimately results in true perspective taking in the sense that the agent's representation of objects after taking the other's perspective roughly matches the second agent's own representation of these objects, allowing the agent to truly see the world as the other does. In the end, this provides a more natural and human-like interaction with the second agent, since the cognitive agent responds as a human plausibly would instead of introducing into the conversation an item (here, a wrench), that the second agent might not even knows exists.

Acknowledgments

This work was partially supported by the Office of Naval Research to Greg Trafton and Alan Schultz under work order number N0001402WX20374. Additional funding was supplied by DARPA IPTO under the MARS program.

References

Anderson, J. R., & Lebiere, C. (1998). *Atomic components of thought*. Mahwah, NJ: Erlbaum.

Carson-Radvansky, L. A., & Logan, G. D. (1997). The influence of functional relations on spatial template construction. *Journal of Memory & Language, 37*, 411-437.

Carson-Radvansky, L. A., & Radvansky, G. A. (1996). The influence of functional relations on spatial term selection. *Psychological Science, 7*, 56-60.

Harrison, A. M., & Schunn, C. D. (2003). ACT-R/S: Look Ma, No "Cognitive-map"! In *International Conference on Cognitive Modeling*.

Levelt, W. J. M. (1984). Some perceptual limitations on talking about space. In A. J. van Doorn, W. A. van der Grind & J. J. Koenderink (Eds.), *Limits in perception* (pp. 323-358). Utrecht: VNU Science Press.

Trafton, J. G., Schultz, A. C., Cassimatis, N. L., Hiatt, L. M., Perzanowski, D., Brock, D. P., et al. (under review). Using similar representations to improve human-robot interaction.

Trafton, J. G., Schultz, A. C., Perzanowski, D., Adams, W., Bugajska, M. D., Cassimatis, N. L., et al. (under review). Children and robots learning to play hide and seek.

Wang, R. F., & Spelke, E. S. (2002). Human spatial representation: Insights from animals. *Trends in Cognitive Sciences, 6*(9), 376-382.