Perspective-taking behavior as the probabilistic weighing of multiple domains

Daphna Heller [a,*], Christopher Parisien [b], Suzanne Stevenson [a]

[a] University of Toronto, Canada
[b] Nuance Inc., Canada

ARTICLE INFO

ABSTRACT

Our starting point is the apparently-contradictory results in the psycholinguistic literature regarding whether, when interpreting a definite referring expressions, listeners process relative to the common ground from the earliest moments of processing. We propose that referring expressions are not interpreted relative solely to the common ground or solely to one's Private (or egocentric) knowledge, but rather reflect the *simultaneous integration of the two perspectives*. We implement this proposal in a Bayesian model of reference resolution, focusing on the model's predictions for two prior studies: Keysar, Barr, Balin, and Brauner (2000) and Heller, Grodner and Tanenhaus (2008). We test the model's predictions in a visual-world eye-tracking experiment, demonstrating that the original results cannot simply be attributed to different perspective-taking strategies, and showing how they can arise from the same perspective-taking behavior.

© 2015 Elsevier B.V. All rights reserved.

Conversation takes place under knowledge mismatch: the conversational partners naturally come to the conversation with different knowledge and beliefs. At the same time, keeping track of what information is or is not shared with one's interlocutor is crucial for conducting felicitous conversation. For example, assertions typically contain information that is not already shared, and, similarly, questions normally ask about information that is not shared (Stalnaker, 1978). Thus, conversational partners must keep track of the shared information (i.e., the *common ground*) alongside maintaining their own private knowledge. To make an assertion, a speaker has to use both types of information simultaneously: coming up with the content of an assertion requires using one's own privileged knowledge, whereas determining whether the assertion will be felicitous at the current state of the conversation requires consulting the common ground.

For definite referring expressions (e.g., *the candle*), there has been a debate regarding whether they are interpreted relative to the private (or egocentric) perspective or relative to the common ground. Theoretical approaches suggest that definite reference depends on shared information (e.g., Clark & Marshall, 1981). For example, a felicitous use of *the candle* depends on there being a

uniquely-identifiable candle in the common ground (Gundel, Hedberg, & Zacharski, 1993). However, in an influential study, Keysar, Barr, Balin, and Brauner (2000) showed that listeners do not, in fact, restrict their attention to objects in common ground. They take this result as indicating that listeners initially interpret definite referring expressions relative to their private (or egocentric) perspective, integrating common ground information only if this initial processing leads to failed reference (see also Keysar, Barr, Balin, & Paek, 1998; Keysar, Lin, & Barr, 2003). Other work, starting with Nadig and Sedivy (2002), has shown instead that listeners are not initially egocentric, but rather show sensitivity to the common ground perspective from the earliest moments of processing (Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Hanna & Tanenhaus, 2004; Hanna, Tanenhaus, & Trueswell, 2003; Heller, Grodner, & Tanenhaus, 2008).

The literature on perspective taking has aimed to reconcile these apparently-contradictory results by pointing to causes that could have led listeners in the different studies to adopt different perspective-taking strategies (Bezuidenhout, 2013; Brown-Schmidt & Hanna, 2011; Hanna et al., 2003; Kuhlen & Brennan, 2013). These explanations assume that listeners choose a single perspective-taking strategy in response to situational factors, and interpret a definite noun phrase either relative to their egocentric perspective or relative to the common ground. The current paper takes a radically-different approach. We propose

---

* Corresponding author at: Department of Linguistics, University of Toronto, 100 St. George Street, Toronto, ON M5S 3G3, Canada.

E-mail address: daphna.heller@utoronto.ca (D. Heller).

that taking perspective entails the *simultaneous use of both the egocentric perspective and the common ground*. We then demonstrate how this new approach can give a unified account of two sets of results that previously have been taken to arise from different underlying mechanisms. Before turning to our proposal, we review the relevant literature.

1. Introduction: Common ground and domains of reference

It has been widely accepted since the work of Russell (1905) that definite descriptions, like *the triangle* or *the white candle*, are used to identify a referent that is *unique* in satisfying the descriptive content of the definite, such as a unique triangle or a unique white candle (for one prominent theory, see Gundel et al., 1993). Uniqueness, however, is not absolute (there is clearly more than one white candle in the world), but is relativized to a contextually-restricted set, known as the *domain of reference* (Roberts, 2003). The domain of reference (or *referential domain*) is not explicitly given as part of the linguistic signal, and thus must be inferred from indirect situational cues: what objects are available in the physical surroundings, what has been said in prior conversation, and also from general world knowledge. Furthermore, domains of reference are not static, but change over time as information is updated over the course of a conversation.

Indeed, a growing body of psycholinguistic evidence shows that listeners can quickly adapt domains of reference as language unfolds. Experiments have revealed that listeners use information in the linguistic signal, like the selectional restrictions of a verb, to restrict their attention to relevant entities. For example, listeners focus on edible things after hearing the verb *eat*, but not after hearing *move* (Altmann & Kamide, 1999). In addition, listeners adapt referential domains using non-linguistic information, such as the affordances of the objects in the physical context (Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Chambers, Tanenhaus, & Magnuson, 2004). For example, Chambers et al. (2004) demonstrate that when listeners interpret an instruction like *Pour the egg in the bowl over the flour*, they develop expectations about whether the noun *egg* will be followed by a modifier (e.g., *in the bowl*) depending on how many eggs in the context are in liquid form and can be thus plausibly be poured.

Despite the remarkable ability to quickly adapt referential domains using linguistic and non-linguistic information, some psycholinguistic findings, such as Keysar et al. (2000), suggest that information about common ground is *not* used in initially restricting referential domains. This is surprising given Clark and Marshall's (1981) theoretical proposal that definite descriptions are used to refer to objects in common ground. Specifically, Keysar et al. (2000) used a referential communication task in which a lab confederate instructed participants to manipulate real-world objects as their eye movements were recorded. Shared information, or common ground, was established by physical co-presence: the objects were placed in a vertical display of cubbyholes, with objects visible to both interlocutors assumed to be in common ground, and objects that were blocked from the confederate's view with a barrier assumed to be in the listener's privileged ground. On critical trials the display contained, for example, three candles that contrasted in size, the largest of which was privileged to the listener, and the confederate speaker instructed the listener to *Pick up the big candle*.[1] The Triplet-Privileged display in Fig. 1 illustrates this situation (except the original display had more distractor objects). Keysar et al. reasoned that if listeners restrict

the domain of reference to common ground, they would not consider the privileged biggest candle as a potential referent. They compared this situation to a control display where the privileged object was unrelated (e.g., an apple). Unlike in the control display, with the critical display they found that listeners *did* look at the privileged object (i.e., the biggest candle), and sometimes even reached for it and touched it. Keysar et al. (2000) interpreted this result as indicating that listeners process initially from their *egocentric* (or *private*) *perspective*, ignoring information about what is in common ground (see also Keysar et al., 2003).

Hanna et al. (2003) proposed an alternative reason for Keysar et al. (2000) not finding an early effect of common ground: It is not that listeners are egocentric, but rather that, in Keysar et al.'s (2000) setup, the privileged object was always a better *perceptual match* to the descriptive content of the referring expression than any of the objects in common ground (in the example above, because it's the biggest candle visible). To test this claim, Hanna et al. (2003, Experiment 1) examined a situation where the privileged object, a red triangle, was identical to the intended referent in common ground, a second red triangle. (Here shared status was established by linguistic mention, not physical co-presence.) Upon hearing an instruction like *Put the blue circle above the red triangle*, listeners were more likely to look at the red triangle in common ground than at the privileged red triangle, and were also faster to choose it, as compared with displays that contained two red triangles in common ground (see Nadig & Sedivy, 2002 for a similar result with young children). This result demonstrates – counter to Keysar et al.'s (2000) claim – that listeners *do* use common ground information from the earliest moments of processing.

Hanna et al. (2003) adopt a constraint-based approach where interpretation is guided by multiple constraints that reflect continuous integration of evidence from multiple sources. Specifically, they propose two probabilistic constraints that can account for both their own results and those of Keysar et al. (2000). First, *the common ground constraint* prefers shared referents over privileged referents, with the strength of bias depending on the strength of the probabilistic cues in the situation that indicate what is in common ground. Second, *the perceptual match constraint* biases reference resolution toward an object whose perceptual properties best match the descriptive content of the referring expression (i.e., the noun and its modifiers) – this constraint is evaluated against all the objects perceptually available to the listener. These two constraints were able to account for both sets of results that were available at that time. First, they account for the pattern in Hanna et al. (2003, Experiment 1) because the *common ground constraint* favors the red triangle in common ground over the privileged red triangle; here the *perceptual match constraint* favors neither, as both objects have the same properties and thus match the definite referring expression ("the red triangle") equally well. Importantly, the same two constraints can also account for the pattern in Keysar et al. (2000): while the *common ground constraint* favors the intended referent in common ground, the *perceptual match constraint* strongly favors the privileged object, because it is a better perceptual match to *big candle* (it is the biggest candle visible to the listener).

The same two constraints cannot, however, account for a more recent result from Heller et al. (2008). Using a similar setup to Keysar et al. (2000), Heller et al. examined the interpretation of an unfolding instruction such as "*Pick up the big candle*" at the point of processing the size adjective. Their experimental design was built on the finding that upon hearing a size adjective – and even before hearing the noun it modifies – listeners expect reference to an object for which there is a size-contrasting object (Sedivy, Tanenhaus, Chambers, & Carlson, 1999). Heller et al. (2008) examined displays that contained one pair of size-contrasting objects

---

[1] The original example in Keysar et al. (2000) had the *smallest* candle privileged, and the instruction was *pick up the small candle*. The example was changed for ease of exposition.