Is perspective-taking in real-time language comprehension spontaneous or strategic? Evidence from eye-movements and grain of coordination

Yipu Wei[1], Yingjia Wan[2], Michael K. Tanenhaus[3]

[1]Peking University, [2]Chinese Academy of Sciences, Institute of Psychology, [3]University of Rochester

Evidence for immediate perspective-taking in referential communication tasks manipulating physical co-presence comes from studies with small referential domains, typically four or five objects (Heller et al., 2008). Perspective-taking with small referential domains might be strategic rather than spontaneous because the potential referents could be attended to and kept in working memory. We evaluated this hypothesis by (1) using a larger number of objects and (2) preceding the referential task with a non-linguistic grain-of-coordination manipulation that either encouraged participants to focus on a small referential domain, which would allow for strategic perspective-taking, or a larger referential domain, which would make strategic perspective-taking implausible.

Listeners participated a referential communication task in Chinese while their eye movements were monitored (75 Chinese participants). A complex visual context with multiple objects was used. A virtual speaker gave temporally ambiguous instructions with scalar adjectives (*dade* "big" in "big cubic block"). Scalar adjectives assume a contrasting object (a small cube). We manipulated whether the contrasting object (a small triangle) for a competitor object (a big triangle) was in common ground (visible to both speaker and listener) or was occluded and thus in the listener's privileged ground (Fig.1), in which case perspective-taking would allow earlier reference-resolution. Before the referential task, subjects participated in a puzzle-solving task with the same virtual speaker, which manipulated referential domain. Pieces were either confined to a small area (requiring fine-grained coordination) or distributed across spatially distinct regions (requiring only coarse-grained coordination). Grain-of-coordination manipulations facilitate pro-social behavior in non-linguistic tasks (e.g., Wan et al., 2019). If perspective-taking is strategic, participants should immediately consider perspective only in the fine-grained condition.

A growth curve analysis (Mirman et al., 2008) on the proportion of looks to the target and target-set (target and target-contrast) during the scalar adjective region (Fig.2). revealed an effect of ground in both fine and coarse-grained coordination groups: The proportion of looks in the privileged condition is higher than those in the shared condition (target: $\beta = 0.33$, $SE = 0.02$, $z = 16.22$, $p < .001$; target-set: $\beta = 0.42$, $SE = 0.02$, $z = 22.48$, $p < .001$), providing evidence that perspective influences real-time comprehension even in a complex visual setting. Analysis of changes of looks over time shows a significant interaction effect of ground, coordination and time in target-set analysis ($\beta = -0.48$, $SE = 0.11$, $z = -4.51$, $p < .001$), indicating more time-locked interpretation of referents in the fine-grained coordination group.

Results strongly supported spontaneous perspective-taking: Comprehension was less time-locked in the coarse-grained condition, but participants still used perspective information to identify the target referent earlier when the competitor contrast was in privileged ground, even when participants believed instructions were computer-generated. This result provides the first demonstration that coordination on a non-linguistic task can affect subsequent linguistic processing and lays the foundation for potentially fruitful avenues of new research.
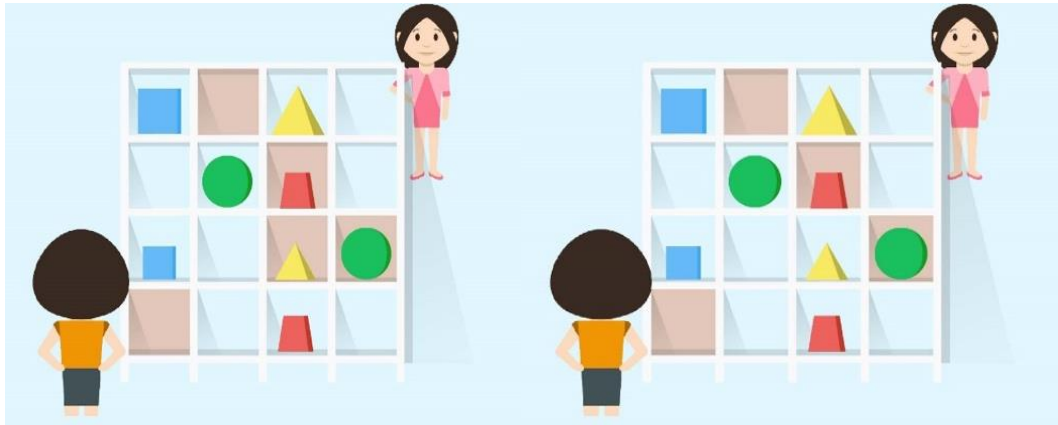
Fig. 1 Example displays of two ground conditions in the online referential communication task (for female participants). Left panel: privileged-ground condition; right panel: shared-ground condition. Four areas of interests were coded for analysis: target (the big cubic block), competitor (the big triangle block), target-contrast (the small cubic block) and competitor-contrast (the small triangle block).
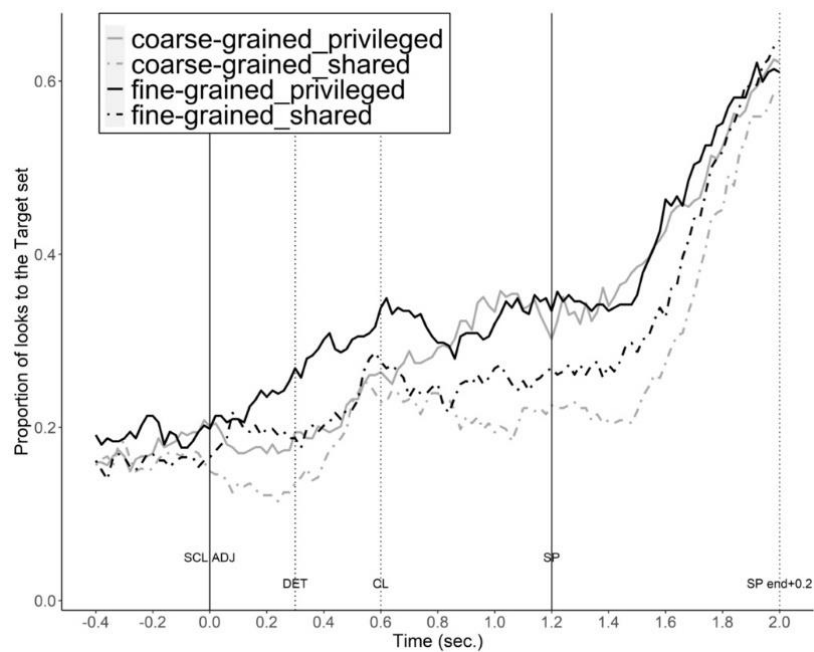

Fig.2 Proportion of fixations on the target-set during the critical time region (from the onset of the scalar adjective to 200ms after the onset of the shape adjective)

Selected references:

Heller, Grodner, & Tanenhaus 2008. *Cognition*, *108*(3).

Keysar, Barr, Balin, & Brauner 2000. *Psychological Science. 11*.

Mirman, Dixon, Magnuson 2008. *Journal of Memory and Language*, *59*(4).

Wan, Fu, & Tanenhaus 2019. *Psychonomic Bulletin & Review*, 26(2).