Title

Effects of Coordination on Perspective-taking: Evidence from Eye-tracking

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 42(0)

Authors

Wei, Yipu
Wan, Yingjia
Tanenhaus, Michael K.

Publication Date

2020

Peer reviewed

Effects of Coordination on Perspective-taking: Evidence from Eye-tracking

Yipu Wei (weiyipu@pku.edu.cn)
School of Chinese as a Second Language, 5 Yiheyuan Road
Peking University, Beijing, 100871, China

Yingjia Wan (wanyj@psych.ac.cn)
CAS Key Laboratory of Behavioral Science, Institute of Psychology, 16 Lincui Road
Department of Psychology, University of Chinese Academy of Sciences
Chinese Academy of Sciences, Beijing, 100101 China

Michael K. Tanenhaus (mtanenha@ur.rochester.edu)
Department of Brain and Cognitive Sciences, 363 Meliora Hall
University of Rochester, Rochester, NY 14627 USA

Abstract

We investigated whether fine-grained coordination in a screen-based puzzle task with a (virtual) partner would influence on-line perspective-taking. Participants played a screen-based puzzle game with a computer player. In the high-coordination condition, the player presented participants with puzzle pieces that could be placed near their partner's last piece. In the low-coordination condition, pieces could only be placed further away from their partner's last piece. Participant's eye movements were then measured in a referential communication task, with the partner giving the instructions, and whether possible competitor referents were in shared or privileged ground. The results demonstrate clear effects of ground and coordination. Participants in both coordination groups were sensitive to the perspective of the interlocutor. In addition, participants in the high-level coordination condition were more sensitive to statistical regularities in the input and their comprehension was more time-locked to the utterance of the speaker.

Keywords: coordination; perspective-taking; joint action; online comprehension; social cognition

1. Introduction

Distinguishing between one's own knowledge and that of an interlocutor, often referred to as "perspective-taking", is central to social cognitive processes, including communication. Moll and Tomasello (2007) argue that perspective-taking skills are cultivated through joint interactions, specifically, interactions with shared intentionality. Indeed, recent developmental studies link collaborative actions with the development of perspective-taking. Cooperative interaction enhances preschoolers' performance on subsequent tasks that require representing the differences between their own desires (Jin, Li, He & Shen, 2017) and visual perceptions (Li et al., 2019), and those of others with whom they are interacting.

The *type* of collaborative action might also be important (Jin, Li, He & Shen, 2018; Wan, Fu & Tanenhaus, 2019). Some collaborative activities emphasize a mutually desired end-product or final state, e.g., a game that requires maximizing the total score of two players, whereas others focus on more continuous coordinated behavior patterns, e.g., partner dancing (Fiebich & Gallagher, 2013). Continuous coordinated experiences are known to promote social bonds, even when the coordination is not intentional. For instance, moving in synchrony increases social closeness, even when participants receive instructions from individual headphones and are not explicitly asked to coordinate (Tarr, Launay & Dunbar, 2016). Since continuous coordination requires participants to pay attention to shared sub-goals (Vesper, Butterfill, et al., 2010), it might have stronger effects on perspective-taking. Indeed, 4-year-olds who closely coordinated with an adult partner were more likely to select an adult-preferred item as a gift for the partner, compared to children who participated in tasks with less coordination (Jin et al., 2018).

Previous work typically measured effortful perspective-taking, that is, asking participants (usually children) to explicitly make judgements about other people's perspectives (Li et al., 2019), which might differ from automatic or spontaneous perspective-taking (Flavell, Everett, Croft & Flavell, 1981; Surtees & Apperly, 2012; Surtees, Apperly & Samson, 2016). If coordination guides perspective-taking, it should continuously influence how people process information. Indeed, people spontaneously represent their partner's point of view in a team game (Surtees et al., 2016).

In this study, we examine whether coordination level in a screen-based puzzle task with a (virtual) partner influences on-line perspective-taking in an unrelated referential communication task. When people converse, their syntactic structures and accents become more similar (Branigan, Pickering, & Cleland, 2000; Giles, Coupland, & Coupland, 1992), their body movements become synchronized (Condon & Ogston, 1971; Shockley, Santana, & Fowler, 2003; Chartrand & Bargh, 1999), and their eye movements become coupled (Richardson, Dale & Kirkham, 2007). The more closely interlocutors coordinate, the better they comprehend each other (e.g. Richardson & Dale, 2005; Shockly, Richardson & Dale, 2009). Drawing on the recent literature on how coordination affects prosocial behavior, we reasoned that the type of a brief coordinative experience with a new interlocutor might influence subsequent perspective-taking during language comprehension.

The ability to distinguish between information that is shared, that is information that is mutually known between

interlocutors, and is thus in "common ground" and information that is privileged to one of the interlocutors (Clark, 1996), can play an important role in resolving the ambiguities that commonly occur in referential expressions. For example, definite reference is used to refer to a uniquely identifiable referent with respect to a circumscribed referential domain. For example, imagine that a speaker says "I had dinner with your daughter last night" to an addressee who has two daughters. The referent of "your daughter" would be ambiguous unless the addressee knew, and was paying attention to, the fact that the speaker was recently at a conference with one of her daughters, and moreover, would have no reason to know that she had another daughter.

Common ground must be inferred using heuristics, including membership in the same community, information acquired in a conversation, and information that is physically co-present to interlocutors. Recent research has focused on the time course with which participants use differences in perspective to resolve referential expressions that would otherwise be ambiguous. Psycholinguists have examined perspective-taking with referential communication tasks that manipulate physical co-presence. For example, an addressee who can see two potential referents for a referring expression might see that the speaker can only see one of the objects. The time course with which the listener uses information about common and privileged ground is assessed by using eye-movements to monitor visual attention during spoken language comprehension (Tanenhaus et al., 1995).

We first manipulated the type of coordinative experience that a participant had with a previously unknown interlocutor. We then monitored eye-movements in a screen-based referential task to examine if, and if so, how, the nature of the experience affected on-line perspective-taking for potential referents that were in shared or privileged ground by virtue of physical co-presence.

Speakers use scalar adjectives, such as "big" in "big candle" when there are two objects of the same type, e.g., two candles that differ (contrast) in size. Building upon research by Sedivy, Tanenhaus et al., (1999), Heller, Grodner & Tanenhaus (2008) found that reference resolution began at the adjective, when there were two big objects, but only one had a size contrast in common ground. Listeners also look more to both the target and its size contrast, which, adopting the terminology introduced by Craig Chambers, we will refer to as a "target-set". The Heller et al. design avoids two problems in earler studies manipulating physical co-presence. First, none of the instructions are either ambiguous or infelicitous. Second, in studies interpreted as evidence that listeners are egocentric (e.g., Keysar, Barr, Balin & Brauner, 2000; Keysar, Lin & Barr, 2003), the privileged ground competitor was a better referential fit for the referring expression than the object in common ground (e.g., "tape" is more commonly used to refer to sticky tape than cassette tape). In the Heller et al. design, privileged and common ground objects are equally good referential fits.

We used a variant of the Heller et al. (2008) task with a screen-based interface. The referring expression was temporarily ambiguous between two potential referents, with either one or both having a size contrast in common ground. The screen-based interface allowed us to control timing in the referential communication task and type of coordinative experience in the puzzle game. One limitation of the Heller et al. study, and others that have found immediate perspective-taking is that the displays contained only four or five objects, which raises the possibility that more egocentric behavior would emerge with displays with more objects. Therefore, we used a display with sixteen grids and eight objects..

In sum, we addressed three questions:

1. Would we find immediate effects of perspective-taking, e.g., would the time course of reference resolution be affected by ground?
2. Would the type of coordinative activity in the puzzle game affect real-time reference resolution?
3. Would type of coordinative activity affect whether or not listeners took ground into account?

Finally, we manipulated whether participants believed they were interacting with another participant or a computer. In the referential communication literature, listeners are sensitive to whether the task is interactive (e.g., Brown-Schmidt & Fraundorf, 2015; Schober & Clark, 1989) and whether the partner is a confederate (Kuhlen & Brennan, 2013). However, research in human-computer interaction show that engaging interactions with robots can produce social effects (Sidner et al, 2005). In the present study, the computer's behavior is designed to mimic that of human's and highly relevant to that of the participant, so the interaction may increase feeling of task sharing and therefore enhance subsequent perspective-taking.

2. Method

2.1 Study Design

We used a 2×2×2 mixed design, with two between-participant variables – coordination level (high vs. low) and partner type (computer vs. human), and one within-participant variable – ground (shared vs. privileged). There was no difference in spent on the puzzle task in the two coordination conditions ($F(1, 71)=0.01$, $p=0.978$). They involve the same amount of work, and achieve the same end product. Task procedures and interfaces were the same, and the participants all played with a computer partner. In the computer partner condition participants were told they were playing with a computer; in the human partner condition participants were told they were playing with another participant.

2.2 Participants

Participants, who were 75 native speakers of Mandarin Chinese from Peking University (mean age=23.15, SD=1.49, 54 females), gave written consent and who were paid for participating. Data from six participants were excluded from analysis because of poor calibration. Participants were

randomly assigned to one of the four between-subject conditions.

2.3 Apparatus

We used an EyeLink-1000 plus eye tracker (SR Research), sampling at 500 Hz. The puzzle task was run by a Python program. Eye-movement data were collected by *Screen-recorder* (version 1.0.0.1264, SR Research). The referential communication task was controlled and recorded by *Experiment-builder* (version 2.2.1, SR Research).

2.4 Procedures

Manipulation Phase - Puzzle Task. Participants played a two-person puzzle game with a computer partner (participants in the human partner condition believed that they were playing with another person). After reading the game instructions, the participant played two practice rounds (a 4-piece puzzle and a 12-piece puzzle). Then the participant and the partner completed the main task with a 48-piece puzzle game (Figure 1).

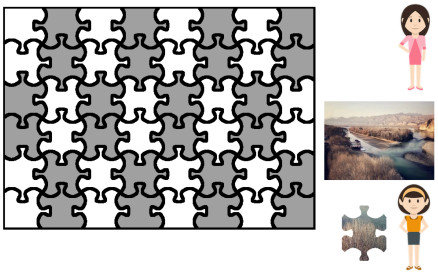


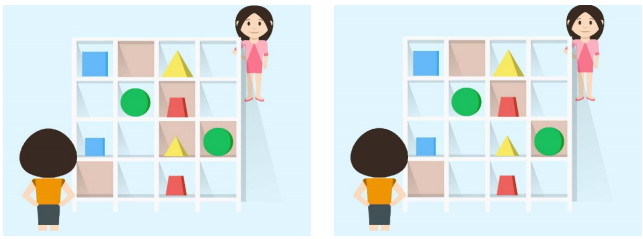
Figure 1: Example display of puzzle task (for female participants).

The interface is illustrated in Figure 1. The participant completed the white areas and the partner completed the grey areas, taking turns placing pieces. The participant first received a piece, displayed beside a same gender avatar. After the participant correctly dragged the piece to the correct place, the next piece for the partner would appear beside the partner's avatar. The participant waited for the partner to place that piece before receiving the next one. This continued until all the pieces were in place.

Pieces were generated by algorithms designed for the two coordination conditions. In the high-coordination condition, puzzle pieces could be placed near their partner's last-placed piece; in the low-coordination condition, pieces could only be placed farther away from each other's last-placed pieces, but close to their own last-placed pieces. Task difficulty was similar. The time the partner spent placing each piece mimicked the time real participants spent placing similar pieces: speed increased as the task progressed, and the partner spent less time on corner and edge pieces.

Test phase: Online referential communication task The perspective-taking task consisted of 16 experimental trials and 16 interspersed filler trials. Each trial paired an auditory sentence and a visual display containing two characters and a shelf with 16 grids (Figure 2).

Participants took the view of the person in front of the shelf (the person in the yellow shirt). The partner was represented by the character on the other side of the shelf (the girl in pink). Grids with the light brown shadow are blocked from the view of the girl behind the shelf. The display in the privileged condition contained five blocked grids; the shared condition contained four blocked grids. There were eight shape objects in both conditions – five were in a shared view in the privileged condition and six in the shared condition.



Privileged-ground condition *Shared-ground condition*

Figure 2: Example displays of two Ground conditions in the online perspective-taking task (for female participants).

After 5s of preview, participants heard pre-recorded instructions in Chinese (voice source: a female native speaker of Chinese) as the following:

Chinese: Qing ba da de na kuai fangxing jimu gei wo.
Gloss translation: Please Ba-construction big MOD DET CL cubic block give me.
English translation: Please give me the big cubic block.

Four areas of interests were coded for analysis: target (the big blue cubic block in Figure 2), competitor (the big yellow triangle block), target-contrast (the small blue cubic block) and competitor-contrast (the small yellow triangle block). In the privileged condition, the competitor-contrast is in the privileged ground of the subject; in the shared condition, the competitor-contrast is in the shared ground. Positions of target and colors and shapes of the target, competitor and target-contrast objects were balanced across grid positions.

After completing both tasks, participants were asked if they noticed anything strange during the experiment. No participant in the human-partner condition suspected the partner was not a real participant.

3. Results and Discussion

We performed two sets of analyses. First we used multilevel logistic regression models to examine target, competitor and target-contrast in the online perspective-taking test across a large analysis region. The critical time window for analysis is from 200ms after the onset of the scalar adjective - *big/small* (1s after the sentence onset and 6s after the start of the picture on the screen) to 200ms after the onset of the shape adjective - *cubic/ sphere/ triangle/ trapezium* (2.2s after the sentence onset and 7.2s after the start of the picture on the screen). During the 1.2s critical time period, a time bin of 20ms was used for analysis. We applied a dummy coding of

eye-movement data: the response variable – *Looks* was coded as '1' if the subject's point of gaze was within a specific interest area during this 20ms time bin, and as '0' elsewise. Multilevel logistic regression models were used to analyze the response in function of *Ground*, *Partner-Type*, *IA (interest area)*, *Coordination* and *Time* (Barr, 2008). The *Time* variable has been centered at 0.6s after onset of the scalar adjective+200ms. Second, we conducted separate analysis on three 600 ms windows linked to theoretically defined regions in the linguistic utterance. These analyses reduce the effects of multiple correlated observations, and provide more detailed information about how critical information in the utterances affected eye-movements.

3.1 Effects of Ground

Figure 3 presents the changes of looks to three interest areas over time: Target, Competitor and Target-contrast. The *Ground* effect was assessed from four aspects: (i) looks to the target object in the privileged-ground condition and the common-ground condition; (ii) whether the *Ground* effect on the looks to the target is interfered by the *Partner-Type*; (iii) comparison between the looks to the Competitor and Target-contrast in the two ground conditions and (iv) looks to the target-set which is composed of the target and target-contrast.

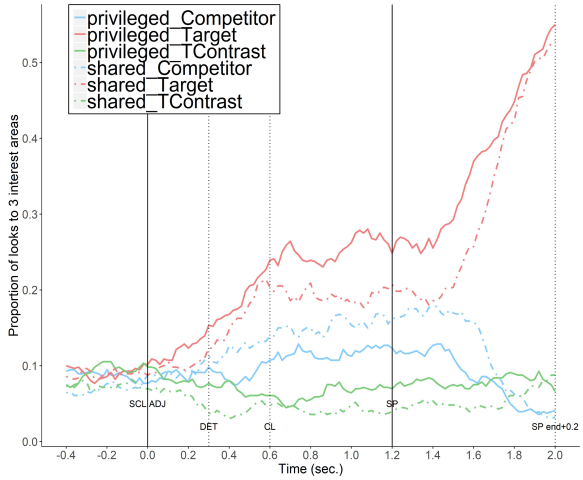


Figure 3: Proportion of looks to the Target, Competitor and Target-contrast in the privileged-ground condition and the shared-ground condition (from 400ms before the onset of the scalar adjective to 200ms after the average offset of the shape (SP) adjective).

3.1.1 Looks to the Target There was a significant difference in looks to the target between privileged-ground condition and shared-ground condition (β=0.330, SE=0.020, z=16.222, p<0.001)[1]. Changes of looks over time also differ, as indicated by an interaction between *Time* and *Ground* (β=0.264, SE=0.058, z=4.583, p<0.001). When the competitor-contrast is in privileged ground, there was a higher and earlier proportion of target looks.

3.1.2 Partner-Type We tested partner type by adding the factor *Partner-type* and the interaction terms with *Time* and *Ground* to the model[2]. *Partner-type* did not affect looks to the target (β=0.016, SE=0.145, z=0.110, p=0.912), or interact with *Time* (β=0.054, SE=0.058, z=0.930, p=0.352). The interaction between *Ground* and *Partner-type* was not significant (β=-0.0767, SE=0.041, z=-1.899, p=0.058). In further analyses, we collapsed across partner-types to increase power.

3.1.3 Competitor and Target-contrast *Ground* should also surface in looks to the target-contrast compared to the size competitor. The final analytical model includes the factor *Ground*, *IA*, *Time*, and interaction terms[3]. There was an interaction of *Ground* and *IA*: in the shared conditions there were *fewer* looks to the target-contrast compared to the competitor (β=-1.351, SE=0.031, z=-43.054, p<0.001). In the privileged-ground conditions there were *more* looks to the target-contrast, in comparison to the shared-ground conditions (β=0.836, SE=0.042, z=19.811, p<0.001).

3.1.4 Looks to the Target-set *Ground* effects were also observed in the looks to the target-set[4]. Proportion of looks to the target-set was significantly higher in the privileged-ground condition (β=0.421, SE=0.019, z=22.482, p<0.001). There was a significant interaction of *Ground* and *Time*: in the privileged-ground condition, looks to the target-set increased much faster as time passed than in the shared-ground condition (β=0.248, SE=0.053, z=4.684, p<0.001).

In sum, participants in both partner groups showed early sensitivity to perspective information. Immediately after hearing the scalar adjective, participants took into account the speaker's perspective. They did not consider a potential referent for an expression beginning with a size adjective when its size contrast was in their privileged ground.

Eye-movements to the target-contrast and competitor provide additional support for early perspective-taking. Although there was a general tendency to look more to the competitor (i.e. the big yellow triangle) as people heard the word *big,* participants in the privileged-ground condition paid more attention to the target-contrast (the small blue cubic) compared to the shared-ground condition.

3.2 Effects of Coordination

We examined effects of coordination level by analyzing the proportion of looks to the target in different *Ground* and *Coordination* conditions.

[1] Model1: glmer((looksattarget==1) ~ ground*timect + (1|PP) + (1|item), data, family=binomial)

[2] Model2: glmer((looksattarget==1) ~ ground*timect + partner_type*timect + partner_type*ground + (1|PP) + (1|item), data, family=binomial)

[3] Model3: glmer((looksattarget==1) ~ ground*IA*timect + (1|PP) + (1|item), data, family=binomial)

[4] Model4:glmer((looktotargetset==1)~ground*timect+(1|pp)+(1|item), data, family=binomial)

3.2.1 Looks to the Target Figure 4 shows the proportion of fixations over time for the four conditions. Proportion of looks to the target within the critical time window were analyzed. Adding the three-way interaction of *Time*, *Ground* and *Coordination* did not change the model fit (χ^2 = 0.361 df=1, p=0.548). Thus, the final model[5] includes the three factors and the two-way interactions.

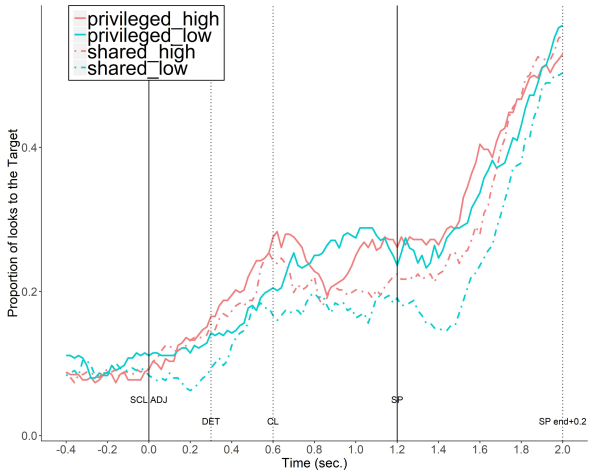


Figure 4: Proportion of looks to the Target in different *Ground* and *Coordination* conditions.

We had initially hypothesized that participants in the high-coordination condition would be more likely to take ground into account more than participants in the low-coordination condition. While there were clear effects of *Coordination* on the time course of processing, as discussed below, participants in both groups showed ground effects.

The *Ground* effects – a higher proportion of looks to the target and a sharper tendency of increased looks to the target over time in the privileged condition (reported in 3.1.1) are modified by *Coordination*. The difference in looks to the target between shared-ground and privileged-ground is smaller in the high-coordination group, compared to the low-coordination group (β=-0.223, SE=0.041, z=-5.494, p<0.001). As shown by Figure 4, looks to the target increase in both privileged/high condition and shared/high condition – suggesting that the high coordination group was developing a strategy of tracking the statistics of where the targets had appeared to predict the likely locations of the next target. Reports from some participants after the experiment supported this interpretation [6].

Despite a tendency of the high-coordination group to fixate more on the target compared to the low-coordination group (β=0.325, SE=0.144, z=2.257, p=0.024), participants in the high-coordination group shifted attention away from the target more quickly than people in the low-coordination group (β=-0.250, SE=0.058, z=-4.341, p<0.001). This is reflected in a drop of proportion of looks to the target in the high coordination groups after it peaks around 0.6s, compared to the low-coordination group, where a drop appears around 1s.

3.2.2 Looks to the Target-set Looks to the target and its size contrast both reflect processes associated with identifying the referent of an expression with scalar contrast. Therefore, we combined looks the target and the target-contrast into a target-set. Figure 5 shows changes in looks to the target-set over time in four conditions.

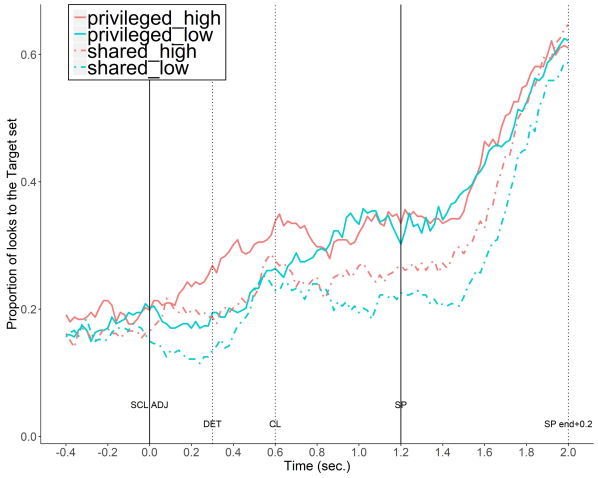


Figure 5: Proportion of looks to the Target-set (Target and Target-contrast) for *Ground* and *Coordination* conditions.

The *Ground* effect on looks to the target-set is consistent with the findings in Section 3.2.4. The effect of *Ground* is significant (β=0.461, SE=0.027, z=17.149, p<0.001), as is the interaction of *Ground* and *Time* (β=0.486, SE=0.076, z=6.404, p<0.001) [7]. In addition, there is a significant interaction effect of *Ground*, *Coordination* and *Time* (β=-0.479, SE=0.106, z=-4.513, p<0.001).

We disentangled the three-way interaction effect with a pair-wise comparison among four conditions (privileged-ground + high-coordination, privileged-ground + low-coordination, shared-ground + high-coordination, shared-ground + low-coordination). In the two shared-ground conditions, changes of looks to the target-set in the high-coordination group are not different from those in the low-coordination group (β=-0.040, SE=0.079, z=-0.514, p=0.608)[8]. However, in the two privileged-ground conditions, the proportion of fixations in the high-coordination group is significantly different from the low-coordination group (β=-0.545, SE=0.074, z=-7.409, p<0.001)[9]. Note that in Figure 5,

[5] Model5: glmer((looksattarget==1) ~ ground * timect + coordination * timect + ground * coordination + (1|PP) + (1|item), data, family=binomial)

[6] In the first four items, the increase of looks to the target in the shared condition is much slower compared to the privileged condition in the high-coordination group (β=-0.312, SE=0.137, z=-2.283, p=0.022)

[7] Model6: glmer((looksattargetset==1) ~ ground * coordination * timect + (1|PP) + (1|item), data, family=binomial)

[8] Model7: glmer((looksattargetset==1) ~ coordination * timect + (1|PP) + (1|item), data, family=binomial)

[9] Model8: glmer((looksattargetset==1) ~ coordination * timect + (1|PP) + (1|item), data4, family=binomial)

the proportion of looks to the target-set in privileged-ground condition diverge from that in the shared-ground condition and peak earlier in the high-coordination group compared to the low-coordination group.

Results from the target and the target-set analyses show an effect of coordination on comprehension. People who had experienced fine-grained coordination showed more time-locked comprehension to the speaker's instructions in the referential communication task. They were also better at tracking the statistics of the target locations. While, both coordination groups used perspective information as soon as they heard the scalar adjective, the high-coordination group resolved the target reference more quickly.

3.3 Window-based analysis

Eye-movements were analyzed for three 600ms time windows. The first window (baseline) captured eye fixations from -400ms before the beginning of the scalar adjective (e.g. *da* 'big') till 200ms after it. The second (early) starts from 200ms after the onset of the scalar adjective and ends at 200ms after the onset of the classifier (e.g. *kuai* 'piece'). The third (late) contained eye fixations from 200ms after the onset of the classifier till 200ms after the onset of the disambiguating shape information (e.g. *fangxing* 'cubic'). We analyzed the proportion of fixations with multilevel linear regression models[10].

Effects of *Ground* emerged in three different windows. The proportion of fixations to the target in the privileged-ground condition were significantly higher than those in the shared-ground condition in both the early and late windows (early: β=0.036, SE=0.015, t=2.469, p=0.014; late: β=0.067, SE=0.018, t=3.793; p<0.001), but not in the baseline window (β=0.031, SE=0.018, t=1.658, p=0.098). *Ground* affected looks to the target-set in the same direction during all three windows: more attention was paid to the target-set under the privileged-ground condition in comparison to the shared-ground condition (baseline: β=0.035, SE=0.017, t=2.067; p=0.039; early: β=0.051, SE=0.017, t=3.001; p=0.003; late: β=0.092, SE=0.019, t=4.841, p<0.001). *Ground* effects were also pronounced when comparing looks to the competitor and those to the target-contrast. During all three windows, there were in general less fixations to the target-contrast than to the competitor (baseline: β=-0.026, SE=0.012, t=-2.091; p=0.037; early: β=-0.072, SE=0.011, t=-6.35; p<0.001; late: β=-0.115, SE=0.012, t=-9.39, p<0.001). However, *Ground* effects interacted with this tendency in the early window and the late window: more looks to the target-contrast were observed in the privileged-ground condition compared to the shared-ground condition (early: β=0.037, SE=0.016, t=2.303; p=0.021; late: β=0.059, SE=0.017, t=3.407, p<0.001), demonstrating a clear influence of *Ground* throughout the referential processing of scalar adjectives.

The influence of *Coordination* varied across different windows. In the baseline window, the high-coordination group fixated more on the target-contrast compared to the low-coordination group (β=0.038, SE=0.018, t=2.179, p=0.029). Looks to the target-set were not significantly different between the two *Partner-type* conditions for the low-coordination group (β=-0.023, SE=0.025, t=-0.897, p=0.373). However, high-coordination participants looked more at the target-set if they believed that they were playing with a real-person partner (β=0.101, SE=0.036, t=2.794; p=0.007). In the early window, participants from the high-coordination group fixated more on the target-set than the low-coordination group did (β=0.047, SE=0.023, t=2.076, p=0.042).

4. Conclusion and implications

We found immediate ground effects, extending the results of Heller et al. (2008) to screen-based conversations with more complex displays. Coordination level in the puzzle game affected the time-locking of the participant's comprehension with the speaker's utterances. Contrary to our initial expectations, both groups considered the speaker's perspective when locating the target object, and use of ground was not modified by partner-type. One possibility is that the statistical pattern for pre-nominal adjectives, and especially scalar adjectives, might be highly consistent across speakers, making contrast effects more automatic than ground effects linked to other linguistic forms.

Nonetheless, there were clear effects of coordination. Comprehension for participants in the high-coordination condition was more time-locked to the utterance of the speaker, This suggests that participating in fine-grained coordinative tasks might facilitate the success of communicative interactions, especially when time pressure is relevant, a possibility that will be important to examine in future research. These findings provide further evidence that coordination is closely associated with language comprehension and communication (e.g., Richardson et al., 2007). It will be important to explore how coordination affects other social cognitive processes such as joint attention. Since perspective-taking is closely related to prosociality (e.g. Vanish, Carpenter & Tomasello, 2009), some of the prosocial effect of coordination might be due to more fluent perspective-taking.

Finally, perspective-taking and the effects of coordination were similar regardless of whether participants believed they interacted with a human or a computer. This raises the possibility that human computer interactions could be structured to improve social cognitive skills in children. Future studies could further test the social effects of interacting with computer with more salient manipulations.

[10] The random structure of the models contained the random slopes of subject and item. Random slopes were not included in the analytical models due to failures of model convergency.

References

Barr, D. J. (2008). Analyzing "visual world" eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474.

Branigan, H.P., Pickering, M.J., & Cleland, A.A. (2000). Syntactic coordination in dialogue. *Cognition*, *75,* B13–B25.

Brown-Schmidt, S. & Fraundorf, S. (2015). Interpretation of informational questions modulated by joint knowledge and intonational contours. *Journal of Memory and Language, 84*, 49-74.

Chartrand, T.L., & Bargh, J.A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*, 893–910.

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Condon, W., & Ogston, W. (1971). Speech and body motion synchrony of the speaker–hearer. In D. Horton & J.J. Jenkins (Eds.), *The perception of language*. Columbus, OH: Charles E. Merrill.

Fiebich, A., & Gallagher, S. (2013). Joint attention in joint action. *Philosophical Psychology*, *26*(4),571–587.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young childrens knowledge about visual-perception – Further evidence for the level 1-level 2 distinction. *Developmental Psychology*, *17*, 99–103.

Giles, H., Coupland, N., & Coupland, J. (1992). Accommodation theory: Communication, context and consequences. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation*. Cambridge, England: Cambridge University Press.

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, *108*(3), 831–836.

Jin, X., Li, P., He, J., & Shen, M. (2017). Cooperation, but not competition, improves 4-year-old children's reasoning about others' diverse desires. *Journal of Experimental Child Psychology*, 157, 81–94.

Jin, X., Li, P., He, J., & Shen, M. (2018). How you act matters: The impact of coordination on 4-year-old children's reasoning about diverse desires. *Journal of Experimental Child Psychology*, *176*, 13–25.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32–38.

Keysar, Boaz, Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25–41.

Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic bulletin & review, 20*(1), 54-72.

Li, P., Jin, X., Liao, Y., Li, Y., Shen, M., & He, J. (2019). Cooperation turns preschoolers into flexible perspective takers. *Cognitive Development*, *52*, 100823.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: *Growth curves and individual differences. Journal of Memory and Language, 59*(4), 475–494.

Moll, H., & Tomasello, M. (2007). Cooperation and human cognition: The Vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362*, 639–648.

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, *29*(6), 1045–1060.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination: Common ground and the coupling of eye movements during dialogue. *Psychological Science*, *18*(5), 407–413.

Sedivy, J. C., K. Tanenhaus, M., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive psychology, 21*(2), 211-232.

Shockley, K., Richardson, D., & Dale, R. (2009). Conversation and Coordinative Structures. *Topics In Cognitive Science*, *1*(2), 305-319.

Shockley, K., Santana, M.V., & Fowler, C.A. (2003). Mutual interper- sonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 326–332.

Sidner, C. L., Lee, C., Kidd, C., Lesh, N., & Rich, C. (2005). Explorations in engagement for humans and robots. *arXiv preprint cs/0507056*.

Surtees, A. D., & Apperly, I. A. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Developmen*t, *83*(2), 452–460.

Surtees, A., Apperly, I., & Samson, D. (2016). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition*, *150*, 43–52.

Tanenhaus, M. K., Spivey-knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Tarr, B., Launay, J., & Dunbar, R. I. (2016). Silent disco: dancing in synchrony leads to elevated pain thresholds and social closeness. *Evolution and Human Behavior*, *37*(5), 343-349.

Vaish, A., Carpenter, M., & Tomasello, M. (2009). Sympathy through affective perspective taking and its relation to prosocial behavior in toddlers. *Developmental Psychology*, *45*(2), 534-543.

Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks, 23*(8/9), 998–1003.

Wan, Y., Fu, H., & Tanenhaus, M. K. (2019). Effects of coordination and gender on prosocial behavior in 4-year-old Chinese children. *Psychonomic Bulletin & Review*, *26*, 685–692.