Are you thinking what I'm thinking? Perspective-taking in a language game

Johanne Nedergaard (nedergaard@cc.au.dk)
School of Communication and Culture, Jens Chr. Skous Vej 2
Aarhus University, Denmark

Kenny Smith (Kenny.Smith@ed.ac.uk)
Centre for Language Evolution, 3 Charles Street
University of Edinburgh, United Kingdom

Abstract

Many theories of communication claim that perspective-taking is a fundamental component of the successful design of utterances for a specific audience. We investigated perspective-taking in a constrained communication situation: Participants played a word guessing game where each trial required them to communicate a target word without context. In each game, pairs of participants took turns giving and receiving clues to guess target words, both receiving feedback after each trial. In Experiment 1, none of the measures of participants' performance improved over rounds, suggesting either that participants were unable to improve their perspective-taking or that the task was simply too demanding for other reasons. In Experiment 2, we tested whether this lack of improvement was due to overall difficulty rather than inability to take perspective. While the success rate in Experiment 2 did improve over the course of the game, our analyses indicated that the improvement was due to participants discovering a frequency heuristic (using rarer clue words) rather than improved perspective-taking per se. The results of these two experiments show that improving perspective-taking adaptively is very difficult when there is no context to ground either signal choice or interpretation.

Keywords: communication; perspective-taking; audience design; interaction; word associations

Introduction

One of the most immediately relevant mechanisms underlying the design and interpretation of utterances is perspective-taking; what people do when they take the knowledge and beliefs of their interlocutors into account. Some theories claim that this mechanism is key (see e.g. Sperber & Wilson, 1995; Clark, 1996) while others argue that humans communicate successfully by gradually aligning their perspectives through repeated interactions (see e.g. Pickering & Garrod, 2004; Barr & Keysar, 2004). The empirical evidence is mixed (see e.g. Barr & Keysar, 2006): Some studies show evidence of perspective-taking while others show a somewhat surprising egocentricity. The present study aims to explore whether and how people are able to use perspective-taking adaptively to increase success in a constrained communication situation through modeling their interlocutor's semantic network.

Perspective-taking as a Stand-alone Mechanism

Sulik and Lupyan (2018a) tested perspective-taking in novel signaling tasks in a series of experiments where participants had to provide a clue word to make their partner guess a target word. To achieve success on the task, directors needed to provide a clue word that would strongly trigger the target word – thus, they had to search for eligible clue words and simulate listener responses to them in order to identify effective clue words. Sulik and Lupyan manipulated whether speaker and hearer had symmetric or asymmetric perspectives in this task. A symmetric trial occurred if the first associate of the target word also had the target word as its first associate. For example, most people say 'night' when cued with 'day', and most people say 'day' when cued with 'night'. In this type of trial, the director could succeed by just providing her own first association as a clue word.[1] An example of an asymmetric trial would be with the target word 'dolphin'; in response to 'dolphin', most people say 'mammal' but in response to 'mammal', most people do not say 'dolphin' – they say 'animal'. Only success on asymmetric trials would be evidence of perspective-taking as this is the only case where the director using her own first association as a clue is highly unlikely to lead to a successful guess.

Participants in Sulik and Lupyan's (2018) study generally failed to take perspective, instead tending to provide the clue most strongly associated with the target from their own perspective instead of from their partner's perspective. Importantly, participants in this study did not interact directly and were not told after each trial whether their clue or guess had been successful, nor what the matcher had guessed or what the target word was in the case of an incorrect trial, meaning that participants could not adapt to their partner or the task over the course of the experiment. In an unpublished follow-up study, Sulik & Lupyan (2018b) found that participants were able to improve their perspective-taking over repeated interactions when they interacted face-to-face and received feedback after each trial. Here, interaction could plausibly have played a role in explaining success, for example if participants learned to visually signal the

[1] For ease of interpretation, we will refer to the speaker as 'she' and the hearer as 'he' throughout.

upcoming use of a specific clue strategy (e.g. homonyms, antonyms, using rare words, hyponyms, etc.).

In many models of perspective-taking found in the existing literature (see again Pickering & Garrod, 2004; Barr & Keysar, 2004), participants improve their communication by accruing common ground over repeated interactions. In the present study, this cannot be the case as the common ground does not as such increase with each trial (given that each trial presents a novel target word). Instead, improvement or adaptive learning is hypothesised to stem primarily from the participants learning to provide more effective clue words, i.e. clue words with higher backward association strength.

The Present Study

We aimed to test whether people could use perspective-taking adaptively when there were no other alternative means of achieving success. In contrast to Sulik and Lupyan (2018a), we provided both participants with feedback (success, the target word, and the guess word) after each trial. We hypothesised that feedback would play an essential part in learning to improve perspective-taking. In contrast to Sulik and Lupyan (2018b), our participants interacted over a longer period of time but were not able to see each other; we did this to prevent participants from sharing any other context than the trial-by-trial clue words and guess words. We were interested in perspective-taking through modeling of semantic networks on its own and hence did not include the face-to-face interaction from Sulik and Lupyan (2018b). Importantly, target words never repeated, which could lead to successful trials based on memory rather than active perspective-taking. The hypothesised improvement over trials would come from participants learning that using clue words that are salient from their own perspective (e.g. providing 'mammal' as a clue for 'dolphin') is unhelpful and that they have to model their partner's semantic associations to achieve success. The only "common ground" our participants had was the semantic associations of the English language. As we have seen above, symmetry (whether speaker's and hearer's perspectives actually differ), salience, and common ground appear to be important factors for perspective-taking to occur. We operationalised symmetry in a similar way to Sulik and Lupyan (2018a) as egocentric and allocentric salience (i.e. salience from the other person's perspective) by word association strength, a measure of how strongly a word cues another word.

Egocentric and Allocentric Salience

The main measure of interest in this study was backward association strength, which served as a measure of perspective-taking. Director backward association strength was the association strength between the clue and the target (how strongly does the clue 'ocean' cue the target 'whale'?), and matcher backward association strength was the association strength between the clue and the guess (how strongly does the guess 'water' cue the clue 'ocean'?). Thus, backward association strength for both director and matcher represented allocentric salience.

To illustrate what perspective-taking would look like in the present setup: The backward association strength between the target 'plague' and the clue 'rat' is 0.01 whereas the backward association strength between the target 'plague' and the clue 'bubonic' is 0.38. This means that giving 'bubonic' as a clue is more likely than 'rat' to elicit the correct response 'plague'. If the speaker chooses 'bubonic' over 'rat', she is taking perspective. If her clue had been driven by forward association strength, she would have given 'death', the highest ranked forward associate of 'plague' (strength = 0.13). If the director gave 'rat' as a clue, the matcher would be showing perspective-taking if he did not say 'mouse' – the highest forward associate of 'rat' (strength = 0.13) – but instead 'rodent', a target word more likely to have elicited the clue word 'rat' (strength = 0.27). Thus, if perspective-taking improved over rounds, we expected backward association strength to increase and forward association strength to decrease.

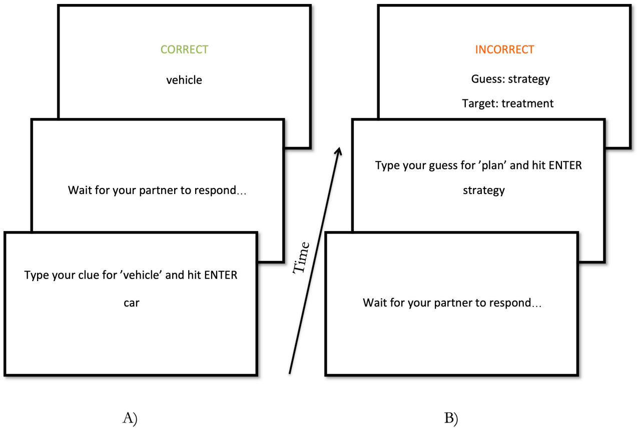
Method: Experiment 1

Participants played a word guessing game in pairs; on each trial, one participant (the director) was required to help their partner (the matcher) guess a single target word by providing only a single clue word, with the roles of director and matcher alternating each trial.

Participants. We recruited 40 participants (10 male and 30 female, mean age = 23.54, range = 18-32) playing as 20 dyads. The participants were recruited from the student population at the University of Edinburgh. All participants were required to be native English speakers above the age of 18 and received £10 for their participation. The participants in pairs did not know each other beforehand.

Materials. We selected 120 target words from the most common English nouns (using the iWeb Corpus; Davies, 2018) to ensure familiarity with meaning and spelling. We selected target words such that half the words in each list had top 1 or top 3 symmetric associates, and the other half had asymmetric associates. For example, the target word 'term' is top 1 symmetric because its top associate is 'semester', and the top associate of 'semester' is 'term'; the target word 'vehicle' is top 3 symmetric because one of its top 3 associates ('car') has 'vehicle' as one of its top 3 associates. In contrast, 'project' is an asymmetric target word because none of its top 3 associates ('work', 'task, and 'school') in turn cue 'project' as one of their top 3 associates. All association strength measures came from the large-scale word association study The Small World of Words (SWOW; De Deyne et al., 2018). Sulik and Lupyan (2018a) compared results using the SWOW with results using the University of South Florida (USF) Free Association Norms (Nelson et al., 2004) and the Edinburgh Associative Thesaurus (EAT; Kiss, Armstrong, Milroy, & Piper, 1973) but found that different association norms produced similar results. Therefore, we only used the SWOW as it was the more extensive database.

Procedure. Participants were told they would be playing a word guessing game where the aim was to help their partner guess the target word using only one clue word. Participants were seated in separate booths and communicated over networked computers using custom-written software in PsychoPy (Peirce et al., 2019). Participants took turns sending and receiving clues and both received feedback after each trial, being given the target, the guess, and whether the guess was correct or incorrect (see Figure 1 for example trials in Experiment 1).



CORRECT

vehicle

Wait for your partner to respond...

Type your clue for 'vehicle' and hit ENTER

car

INCORRECT

Guess: strategy

Target: treatment

Type your guess for 'plan' and hit ENTER

strategy

Wait for your partner to respond...

Figure 1. *(A) a successful director trial: the director provides the clue 'car' for the target 'vehicle'. (B) an unsuccessful matcher trial: the matcher guesses 'strategy' from the clue 'plan'.*

The director on a given trial could type in any real English clue word (except one identical to the target word) and there was no time limit on giving a response. The matcher received the clue word and typed in their guess, again without a time limit. Dyads had 20 trials in each round and played six rounds; the entire experiment lasted approximately one hour.

Results: Experiment 1

If participants are able to perspective-take adaptively, we expected both success rate (proportion of correct guesses by the matcher) and perspective-taking to increase over the course of the game. The primary dependent variables were the binary success outcome (match between target and matcher's guess) and director backward association strength (with higher values indicating clue words that more strongly cued the target, i.e. suggesting better perspective-taking by the director). The secondary dependent variables were director and matcher forward association strength, matcher backward association strength, and the word frequency of the clue word (Zipf value; Van Heuven, Mandera, Keuleers, & Brysbaert, 2014). Zipf value of the clue words was included as an additional measure of how the clue words themselves changed over time. These additional measures were secondary in the sense that they could not on their own prove or disprove the hypotheses but could lend support in either direction.

To better assess which properties of the target words influenced success rate, we also examined the independent variables accessibility (1st, 2nd, 3rd and 4th quantiles of director backward association strength of optimal clue word) and symmetry (top 1 symmetric, top 3 symmetric, and asymmetric). The accessibility variable operationalised the existence of 'good' clue words, i.e. words that strongly and specifically cued a given target. For example, the target word 'eye' had a good potential clue word in 'retina' (p(eye|retina) = 0.33) whereas the target word 'department' did not have a particularly good potential clue, the best one being 'bureau' (p(department|bureau) = 0.02). 'Eye' therefore falls in the 4th quantile of accessibility (it should be relatively accessible, in that a good clue word does exist) whereas 'department' falls in 1st quantile of accessibility. In Experiment 1, the maximal backward association strength (backward association strength between target and the optimal clue word) had an overall mean of 0.16 and ranged from 0.01 to 0.33; the 1st quantile was 0.09 and the 3rd quantile was 0.23.

Data cleaning and preparation. Three dyads ran out of time and did not complete all six rounds: One dyad only played three rounds, and two dyads only played four rounds. Their trials were still included in the analyses. Guess words with spelling mistakes, typos, plurals, and other standard spellings counted as correct in both experiments. As participants could type any word for both clues and guesses, some of the clues and guesses did not appear in the SWOW norms and we were therefore unable to score association strengths. This was the case for 387 of the trials for director forward association strength (17.12 %), 516 of the trials for matcher forward association strength (22.83 %), 426 of the trials for director backward association strength (18.85 %), 437 of the trials for matcher backward association strength (19.34 %), and 266 of the trials for Zipf value (11.77 %).

Descriptive statistics. See Figure 2A for average success over rounds and Figure 2B for average director backward association strength over rounds.

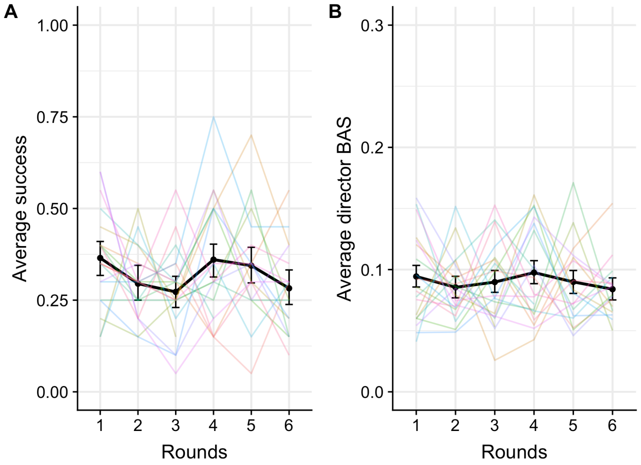


Figure 2. *Success (A) and director backward association strength (B) over rounds. The black lines indicate averages across dyads with error bars showing bootstrapped 95 % confidence intervals. Coloured lines represent individual dyads. Note the different scales in A and B.*

Success. We analysed success predicted by round number with a binomial mixed-effects regression model with correlated random slope and intercept for target word and uncorrelated random slope and intercept for dyad. This model indicated no effect of round on success (p = .711).[2]

Director backward association strength. A linear mixed-effects regression model indicated no effect of round on director backward association strength (p = .193).

Further measures of improvement. None of the other dependent variables (Zipf value, director and matcher forward association strength, and matcher backward association strength) showed significant improvement across rounds (all p > .13).

Symmetry. A binomial mixed-effects regression model with success predicted by symmetry indicated that top 1 symmetric target words were significantly easier to communicate than asymmetric targets words (β = 1.10, SE = 0.29, z = 3.82, p < .001). Top 3 symmetric targets were not significantly easier compared with asymmetric targets (p = .202).

Accessibility. Accessibility was also a significant predictor of success with 3rd and 4th quantiles being easier to communicate than the baseline 1st quantile (3rd: β = 1.26, SE = 0.24, z = 5.18, p < .001; 4th: β = 1.84, SE = 0.24, z = 7.59, p < .001).

Discussion: Experiment 1

Symmetry and accessibility predicted success in the language game, but participants' perspective-taking did not improve over time. Top 1 symmetric items were significantly easier to communicate successfully than top 3 symmetric or asymmetric targets, indicating that directors found it difficult to take the matcher's perspective. Target words with potential clue words that belonged to the top half of maximal backward association strength were also significantly easier for directors to convey and matchers to guess, suggesting that directors were sensitive to the existence of clue words. The results from Experiment 1 indicate that people cannot learn to improve their perspective-taking, even when they get feedback. This is somewhat surprising, given that participants in Sulik and Lupyan (2018b)'s study improved over time with the same level of feedback as in the present experiment. It is still possible, however, that participants relied on perspective-taking for the success they did have, but that the search for potential clue words and simulations of guesses were too demanding for them to improve over time. To test whether the lack of improvement was due to lack of perspective-taking or perhaps a too-challenging task, in Experiment 2 we test whether performance could improve with richer feedback and more accessible targets.

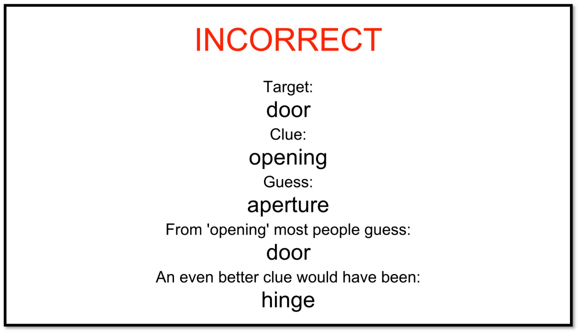
Method: Experiment 2

Participants. For Experiment 2, we again recruited 40 participants playing as 20 dyads (12 male and 28 female, mean age = 23.33, range = 19-52). As in Experiment 1, the participants in pairs did not know each other beforehand.

Materials. We selected 120 new target words from the most strongly cued words in the SWOW dataset in an attempt to make the task less challenging. None of the target words had top 1 symmetric associates but they could have top 3 symmetric associates.

Procedure. Aside from the feedback provided in Experiment 1, participants were additionally after unsuccessful trials told what the optimal clue word would have been, and the top associate in response to the clue the director actually gave (see Figure 3 for an example of the changed feedback screen in Experiment 2).

[2] For following models, we adopted the following procedure: We attempted to model the maximal structure suitable for the experimental design, i.e. random slopes and intercepts for dyads or participants, varying by the dependent variable, and target words. For dependent variables that were determined by the dyad (like success), we used dyad for the random effects – for the dependent variables that were determined by individual participant (like forward and backward association strength, etc.), we used participant for the random effects. If the model then failed to converge or produced singular fit warnings, we set the slopes and intercepts to be orthogonal. If the problems remained, we excluded random effects based on conceptual reasoning (i.e. for predicting success across rounds, we wanted to control for the slope of the dyads so prioritised keeping the random slope over intercept for dyads). In the text, we report the most complex model that we were able to fit. Additionally, we checked the random slopes and correlations between random effects for magnitude and that the unconverged models yielded similar results to the final models, which was the case for all of them.



INCORRECT

Target:
door
Clue:
opening
Guess:
aperture
From 'opening' most people guess:
door
An even better clue would have been:
hinge

Figure 3. *An example of the feedback screen after an incorrect trial in Experiment 2.*

Results: Experiment 2

The dependent and independent variables were identical to Experiment 1, excluding the symmetry variable which was unnecessary as all target words in Experiment 2 had asymmetric associates. In Experiment 2 where all the target words were selected to have good potential clue words, the maximal backward association strength had an overall mean of 0.30 and ranged from 0.23 to 0.38. The 1st quantile was 0.28 and the 3rd quantile was 0.31.

Data cleaning and preparation. One dyad only played three rounds, and two dyads only played five rounds. Their trials were still included in the analyses. Three trials in total were excluded from the analyses due to technical errors while recording responses. As in Experiment 1, we were not able to look up information for all the clue and guess words. This was the case for 325 of the trials for director forward association strength (14.21 %), 425 of the trials for matcher forward association strength (18.58 %), 380 of the trials for director backward association strength (16.62 %), 408 of the trials for matcher backward association strength (17.84 % of the trials), and 285 of the trials for Zipf value (12.46 %).

Descriptive statistics. See Figure 4A for average success over rounds and Figure 4B for average director backward association strength over rounds.

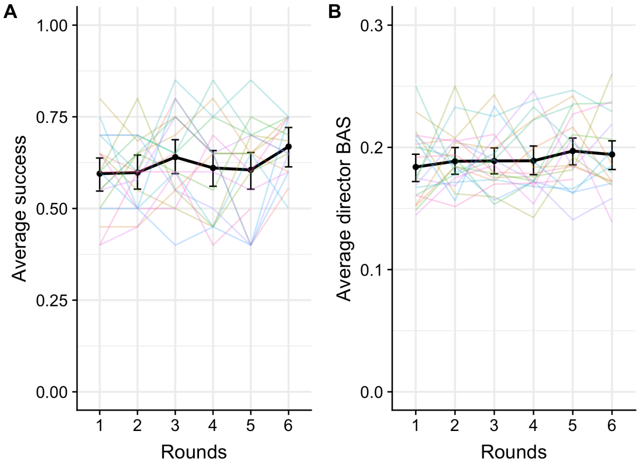


Figure 4. *Success (A) and director backward association strength (B) over rounds. Plotting conventions as in Figure 2. Note the different scales in A and B.*

Success. A binomial mixed-effects regression model indicated a marginally significant positive effect of round on success (β = 0.07, SE = 0.04, z = 1.75, p = .081).

Director backward association strength. A linear mixed-effects regression model indicated a significant positive effect of round on director backward association strength (β < 0.01; SE < 0.01, t(82.47) = 2.06; p = .043).

Further measures of improvement. The secondary dependent variables matcher forward association strength and Zipf value both appeared to improve significantly over rounds. Director forward association strength showed marginally significant improvements (β < -0.01; SE < 0.01, t(132.6) = -1.87, p = .063). A linear mixed-effects model of Zipf value predicted by round indicated directors produced clue words with lower Zipf value – i.e. lower frequency – with every round (β = -0.04; SE = 0.01, t(46.83) = -3.64; p < .001). A model of matcher forward association strength predicted by round suggested that this measure decreased over rounds (β < -0.01; SE < 0.01, t(1811) = -2.45; p = .014). Matcher backward association strength did not appear to improve over rounds (p = .097). Importantly, when we compared a model with only Zipf value as a predictor of success to a model with both Zipf value and round as predictors, the latter did not show significantly improved fit (χ^2 = 2.60, p = 0.107), meaning that round did not appear to explain additional variance above and beyond that explained by Zipf value. This was also true for the model predicting director backward association strength (χ^2 = 4.11, p = 0.128).

Accessibility. Accessibility was again a significant predictor of success with the 4th quantile being significantly easier to communicate than the baseline 1st quantile (β = 0.99, SE = 0.26, z = 3.79, p < .001).

Discussion: Experiment 2

As in Experiment 1, accessibility was a significant predictor of success, indicating that participants were sensitive to the existence of good and bad clue words. In contrast to Experiment 1, there was some indication that participants were able to slightly improve their performance over the course of the game: success and director forward association strength both showed marginally significant improvement while clue Zipf value, director backward association strength, and matcher forward association strength all increased significantly. However, our model comparisons suggest that directors simply realised that picking less common clue words would improve success.

General Discussion

The results of Experiment 1 indicated participants were unable to spontaneously improve their success rate and perspective-taking in the word guessing game. Experiment 2 provided some evidence that participants were able to adjust to the demands of the game under maximally helpful conditions. Here, participants got more extensive feedback, did not have to switch strategy between symmetric and asymmetric trials, and were shown how to give good clues. Consistent with the previous literature on perspective-taking, symmetry was a significant predictor of success in Experiment 1, supporting the idea that it is more demanding for the director to suppress their own perspective when providing a useful clue. Accessibility was also a significant predictor in both experiments, confirming that target words that had a good clue were easier to communicate than words that had weaker clues. This showed participants were sensitive to the existence of good and bad clue words throughout. It is also important to note that the participants were not completely unable to succeed – getting on average around a third of the words right in Experiment 1 and around two thirds of the words right in Experiment 2 – but that they appeared unable to *improve.*

The question remains what caused the apparent (but slight) improvement over the course of the game in Experiment 2 but not in Experiment 1. There are two plausible explanations: 1) People do use perspective-taking in normal communication but it is too difficult to deploy it in this context-poor setup (which is why we saw it improve somewhat in Experiment 2 which was designed to be helpful), or 2) People do not rely on perspective-taking (in our task, and perhaps in normal communication), and the reason we saw improvement in Experiment 2 was that participants learned to use clue word frequency as a heuristic for providing useful clue words. It seems plausible that directors used frequency as a heuristic for narrowing down the search space (of all the potential clues that come to mind, which is the rarest word?). Most of the optimal clue words that were given as part of the feedback after an incorrect trial were rare (words like 'origami', 'eczema', 'retina', and 'bubonic'), which participants may have picked up on. Our analyses supported the latter explanation as round progression did not explain additional variance in either success or director backward association strength once Zipf value was included as a predictor.

Taken together, the results of the two experiments appear to show that active perspective-taking is cognitively demanding and that the extent to which people can deploy it spontaneously in audience design, at least in this challenging guessing-game task, is limited.

Conclusion

Even if it is somewhat unclear what exactly caused the shift in results from Experiment 1 to Experiment 2, we can conclude that improving perspective-taking is effortful and demanding, especially in circumstances without context where the search space for both signal and interpretation are unconstrained. The failure to improve in Experiment 1 was probably located at the search part of the task, as results from Experiment 2 indicated participants were able to improve their performance to some extent when they were given more guidance about the kinds of words they should be searching for. In contrast to most previous studies, the present study examined the power of perspective-taking as a stand-alone mechanism where the building up of shared context is extremely limited and the perspective-taking itself is the only thing that can improve. The findings indicate that while perspective-taking might play a foundational role in ordinary communication, successfully applying perspective-taking requires context or feedback; when these are constrained, perspective-taking breaks down.

References

Barr, D. J., and Keysar, B. (2004). Making Sense of How We Make Sense: The Paradox of Egocentrism in Language Use. In H. L. Colston & A. N. Katz (Eds.). *Figurative language comprehension: Social and cultural influences* (pp. 21-42). New York: Routledge.

Barr, D. J., & Keysar, B. (2006). Perspective taking and the coordination of meaning in language use. In *Handbook of psycholinguistics* (pp. 901-938). Academic Press.

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

Davies, Mark. (2018). The 14 Billion Word iWeb Corpus. Available online at https://corpus.byu.edu/iWeb/

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The "Small World of Words" English word association norms for over 12,000 cue words. *Behavior Research Methods*, *51*(3), 987–1006.

van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1176-1190.

Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In A. Aitken, R. Bailey, & N. Hamilton-Smith (Eds.): *The Computer and Literary Studies*. Edinburgh, UK: Edinburgh University Press.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida word association, rhyme, and word fragment norms. *Behavior Research Methods, 36*(3), 402–407.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior research methods*, *51*(1), 195-203.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(02), 169–225.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (Second ed.). Oxford ; Cambridge, Mass.: Blackwell.

Sulik, J., & Lupyan, G. (2018a). Perspective taking in a novel signaling task: Effects of world knowledge and contextual constraint. *Journal of Experimental Psychology: General*, *147*(11), 1619–1640.

Sulik, J. & Lupyan, G. (2018b). Success in signaling: the effect of feedback to signaler and receiver. In Cuskley, C., Flaherty, M., Little, H., McCrohon, L., Ravignani, A. & Verhoef, T. (Eds.): *The Evolution of Language: Proceedings of the 12th International Conference* (EVOLANGXII).