Inhibitory Control Supports Referential Context Use in Language Production and Comprehension

Alison M. Trude (atrude2@jhmi.edu)
Department of Neurology, Johns Hopkins University
1629 Thames St., Suite 350, Baltimore, MD, 21231 USA

Nazbanou Nozari (nozari@jhu.edu)
Department of Neurology, Department of Cognitive Science, Johns Hopkins University
1629 Thames St., Suite 350, Baltimore, MD, 21231 USA

Abstract

Using referential context in language (e.g., saying "*blue* pen" when two different-colored pens are visible) makes communication efficient. But it is still unclear which general cognitive processes support the use of context in conversation. Research on pragmatic use in language implicates working memory and inhibitory control; however, no studies have shown evidence of a shared cognitive mechanism in both production and comprehension within an individual. The current study asked a) whether referential context use is supported by the same cognitive mechanisms in production and comprehension, b) which processes are implicated, and c) whether the nature of the context itself affects processing. Participants completed a referential communication eye-tracking task in which a disambiguating adjective was either necessary or over-informative, as well as a cognitive test battery. The results implicated inhibitory control in both production and comprehension (although the comprehension results were more variable), suggesting a shared underlying cognitive mechanism across domains.

Keywords: language production; language comprehension; discourse; pragmatics; inhibitory control; working memory

Introduction

The ability to take context into account often facilitates communication in interactive settings. Imagine you are cooking with a friend. A big pepper, a small pepper, and a tomato are sitting on the table. You may know that you would like the big pepper; however, if you just tell your friend to, "hand me the pepper," he will likely be confused as to which pepper you mean. Thus, in order to effectively communicate your intent to him, you would need to use a disambiguating adjective (i.e., "hand me the big pepper"). On the other hand, saying "hand me the big tomato" would be confusing, as it implies to your listener that there is more than one tomato to choose from. This paper investigates how speakers and listeners behave in situations when adding or subtracting an adjective is most appropriate for clear communication given the referential context.

Related to the idea that listeners and speakers will tailor their language to the referential context is the Gricean Maxim of Quantity (Grice, 1975), which specifies that speakers should make their utterances only as informative as is required. Thus, enough information should be provided to distinguish the intended referent from its potential competitors (e.g., the big pepper when two peppers are visible), and providing information that is not necessary (e.g., the big tomato, when there are no other tomatoes in view), should be avoided. Although, ideally, following this maxim would help to make communication maximally efficient, in reality, speakers and listeners and often fail to behave in a completely Gricean manner (e.g., Deutsch & Pechmann, 1982; Sedivy, 2005).

Speakers' and listeners' ability to make their utterances optimally informative, or to appropriately interpret the utterance they are hearing, within the current referential context may depend upon the cognitive demands that this process places upon them. Referential context adaptation (RCA) is a complex process, involving not only language production or comprehension, but also selectively attending to certain objects in one's surroundings, remembering what information has already been introduced into the discourse, or refraining from mentioning irrelevant or confusing information, to name a few. Therefore, general cognitive functions, such as working memory (WM) and inhibitory control (IC), could play an important role in RCA.

The current work investigates whether the ability to take referential context into account is a unified construct in language production and comprehension, and under different linguistic demands. If so, we would expect (a) individuals with better RCA to demonstrate this ability in both production and comprehension, and (b) the same general cognitive operations to drive referential context consideration in comprehension and production, and perhaps in situations with different linguistic demands.

Pragmatic Language Use and Cognitive Abilities

The fact that speakers often fail to observe the Gricean Maxim of Quantity provides evidence that taking referential context into account may be a cognitively demanding process. While research in this field has not previously addressed the relationship between RCA and general cognitive abilities, some work on a related linguistic process, perspective-taking, has attempted to identify the cognitive processes underlying pragmatic language use. Perspective-taking and RCA are similar in that both involve the on-line incorporation of referential context in interactive conversation. Thus, it is possible that a similar set of cognitive mechanisms supports both processes.

In one study of perspective-taking, speakers were tested on their ability to refrain from producing descriptions that were over-informative from the perspective of their listener (e.g., “the big star” when two stars are visible to the speaker, but only one is visible to the listener; Wardlow, 2013). The results showed that this ability was correlated with both WM and IC. In comprehension, Brown-Schmidt (2009) found that participants with greater IC were more likely to take the experimenter’s perspective when interpreting their questions about a display containing some pictures that were visible to only the listener. Lin, Keysar, and Epley (2010) found that comprehenders with higher WM capacity performed better on a similar task.

While these studies all point to a role for general cognitive resources in perspective-taking, an important question remains: Do the same abilities underlie perspective-taking in production and comprehension? The two studies that have investigated perspective taking in both domains within the same individuals found contradicting results: Ryskin, Benjamin, Tullis, & Brown-Schmidt (2015) found WM to be correlated with perspective-taking in production, but none of their executive control measures were correlated with comprehension. On the other hand, Nilsen and Graham (2009) found that IC negatively correlated with egocentric behaviors in a comprehension task in 3-5 year olds; however, none of their executive control measures correlated with production. The question, thus, remains: is the ability to take into account the referential context the same in both domains?

The Current Study

The current study investigates how speakers and listeners adapt their language processing to referential context when observing the same visual display. Specifically, we asked: is the ability to take referential context into account the same in production and comprehension, and under different linguistic and contextual demands? To this end, each participant took turns as both speaker and listener in a referential communication task with a visual world design, and their eye fixations, as well as their utterances (when acting as speakers) were recorded. Two conditions with different contextual demands were created: in the Adj+ condition, the most efficient utterance required the inclusion of an adjective (e.g., “Click on the heart below the **green** gorilla,” Figure 1a). In the Adj− condition, the most efficient utterance was one without an adjective (e.g., “Click on the heart below the gorilla,” Figure 1b). To test which cognitive processes underlie RCA in production and comprehension, each participant also completed a battery of tasks that included three WM and three IC measures. Due to the rich set of both independent and dependent variables in the current design, we adopted a statistical approach that is well-suited to handling this type of data structure: partial least squares path modeling (PLS-PM). This method allows us to look at multiple dependent variables simultaneously. In addition, this method allows us to similarly group WM and IC tasks into latent constructs in order to minimize task-specific effects and avoid the issue of collinearity.

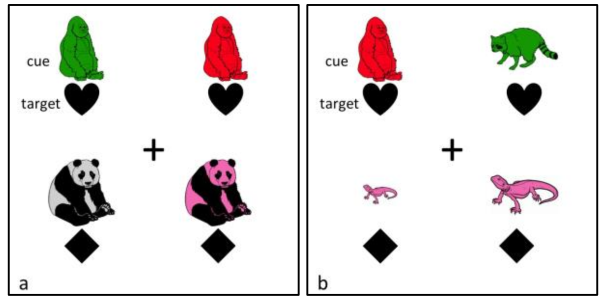


Figure 1: Example Adj+ (a) and Adj− (b) displays used in the referential communication task.

Methods

Participants

Twenty-eight native English speakers, ages 18-30, participated in exchange for $40.

Materials

Referential Communication Task Each display consisted of four black card suit shapes (club, diamond, heart, or spade). Above each shape was a drawing of an animal. The animal stimuli were chosen to be cohort competitors with either a size or color adjective (e.g., big buffalo, green gorilla). On each trial, the target and competitor shapes were the same (e.g., both spades), ensuring that the target shape always needed to be disambiguated from the competitor by describing the animal above the target.

There were two critical trial types. On Adj+ trials (n=96), the same cue animal appeared above the target and competitor shapes in two different colors or sizes. Critically, the animal above the target shape was rendered in its cohort-competitor adjective (e.g., **g**reen **g**orilla). These trials were designed such that, if participants used RCA, in production, they would include the adjective necessary to disambiguate the target cue from the competitor cue, and in comprehension, upon hearing the initial phoneme of the adjective (e.g., the /g/ in green), they would interpret this sound as the beginning of an adjective, and not a noun, and as a result fixate more on the cue above the target (green gorilla) than the cue above the competitor (red gorilla) (Table 1).

On Adj− trials (n=96), different cue animals appeared above the target and competitor shapes. The cue animal above the target shape was always a cohort competitor with the adjective corresponding to the cue above the competitor shape (e.g., red gorilla vs. green raccoon). These trials were designed such that, if participants used RCA, no adjective would be produced in production, as it would be over-informative, and in comprehension, upon hearing the initial phoneme of the target noun (e.g., the /g/ in gorilla), the participant would interpret this phoneme as the onset of a noun (gorilla), and not of the adjective preceding the competitor cue (green raccoon) (Table 1).

Table 1: Action marking good RCA on Adj+ and Adj− trials.

	Adj+	Adj−
Production	Say "the heart under the **green** gorilla."	Say "the heart under the gorilla."
Comprehension	Upon hearing "heart under g…" look for "green," not "gorilla"	Upon hearing "heart under g…" look for "gorilla," not "green"

Cognitive Battery

A spatial WM task (Corsi Block, Kessels et al., 2000) and two linguistic WM tasks (Category and Rhyme Probe, Freedman & Martin, 2001), measuring semantic and phonological WM respectively, were administered. Three IC measures were also employed: Fish Flanker along with embedded NoGo (Nozari, Trueswell, & Thompson-Schill, 2016), as well as semantic blocking elicited through cyclic naming (e.g., Schnur et al., 2006), in which participants named twelve sets of six images, either in semantically-homogenous or heterogeneous blocks. Semantic blocking was determined as the difference in response latencies between the two block types.

Procedure

Participants completed two sessions 3-7 days apart. Each session began with the eye-tracking task, followed by half of the cognitive measures. Stimuli were displayed using MATLAB and Psychophysics Toolbox (Brainard, 1997), and the participant's eye movements during the referential communication tasks were recorded using an Eyelink 1000 Plus desktop-mounted eye-tracker (SR Research).

Referential Communication Task, Session 1. On each trial, a fixation cross was presented for 500 ms before the stimuli appeared. After a 1500 ms preview period, a tone sounded and the target shape began to flash, cuing the participant to begin speaking. After 2.5 seconds, a lower tone sounded, indicating the end of the trial. The participant and experimenter viewed separate monitors containing the same stimuli. The participant instructed the experimenter to click on the target shape that was cued on the participant's screen, using sentences with this structure: "Click on the [target shape] under the [adjective, if needed] [target cue]." They were told that the goal of the task was to provide instructions as quickly as possible, and to avoid unnecessary words to meet the temporal deadline, thus motivating them to drop the adjective when not necessary.

Referential Communication Task, Session 2. Participants followed instructions like the ones they had given during Session 1. Each trial began with a 500 ms fixation point. The stimuli then appeared onscreen, and after a 1000 ms preview period, a pre-recorded instruction played, instructing the participant to click on one of the shapes. Instructions were always pragmatically appropriate (i.e., an adjective was always provided on Adj+ trials, and no adjective was provided on Adj− trials).

Analyses

Behavioral

Participants made a total of 345 RCA errors out of 5376 critical trials (6.4% error rate, *M* = 12.3 errors, *SD* = 8.5). Of these errors, 187 were made on Adj+ trials (i.e., errors of adjective underuse), while 157 were made on Adj− trials (i.e., errors of adjective overuse). Participants were expected to perform near ceiling on the comprehension task, so no accuracy measures were collected.

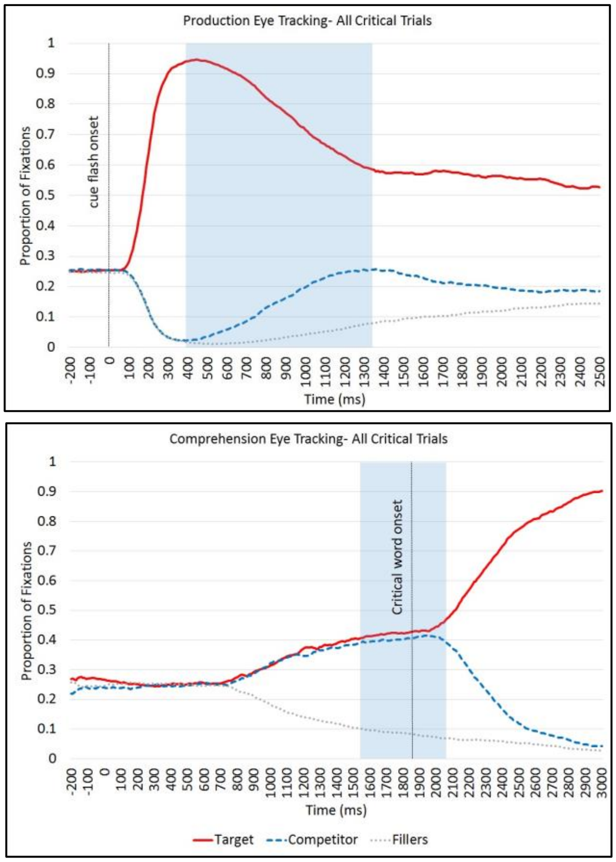
Eye-Tracking

In production, eye-tracking analyses combined each card suit and its animal cue into a single region. Data was analyzed from 400-1350 ms post-target cuing, comprising the time period between the minimum and maximum proportion of fixations to the competitor after attention was initially drawn to the flashing target (see Figure 2a). Competitor fixations were considered critical because, in order to establish the referential context of the display, participants needed to divert their attention from the extrinsically-cued target to its competitor. Thus, our dependent measure of analysis was a competitor advantage score, or the proportion of fixations to the competitor minus the proportion of fixations to the target.

In comprehension, eye-tracking analyses were completed on the target advantage score (proportion of fixations to the target and its cue minus the proportion of fixations to the competitor and its cue) over a time window beginning 300 ms before the onset of the critical word (cue noun in the Adj- condition, and cue adjective in the Adj+ condition) and ending 200 ms after the onset of the critical word (see Figure 2b). The time window was chosen in order to encompass coarticulatory cues from the word preceding the critical word, as well as processing of the initial cohort phoneme of the critical word. This target preference measure should indicate how well participants took referential context into account (i.e., disregarded the competitor). By using target advantage as a DV in comprehension and competitor advantage in production, we ensured that in both cases, a more positive eye-tracking score would index better RCA.

Partial Least Squares Path Modeling (PLS-PM)

Analyses were conducted using partial least squares path modeling, implemented in the plspm package in R (Sanchez, 2013). PLS-PM is a partial least square approach to Structural Equation Modeling suitable for analyzing the relationship between latent variables (psychological constructs such as WM) and manifest variables (observed data from tasks assumed to index these variables) as a network of multiple interconnected linear regressions.



a.
b.

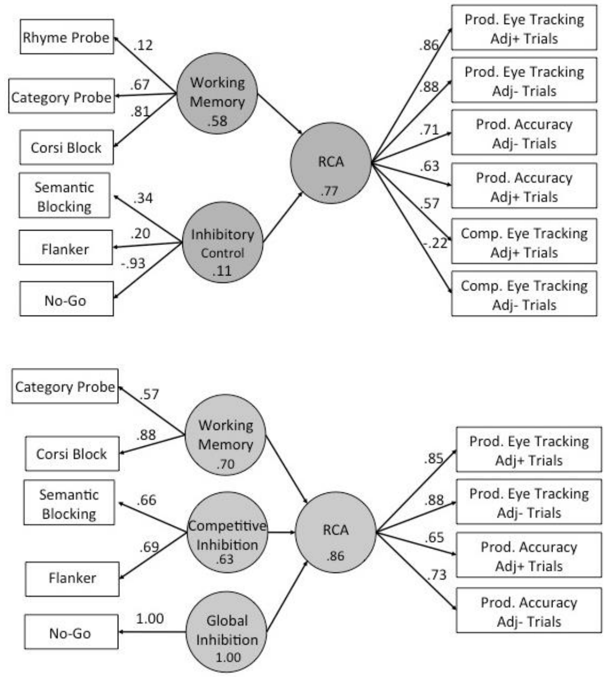
Figure 2: Time course of eye-tracking data for all critical trials in Production (a) and Comprehension (b). Highlighted regions indicate windows of analysis.

Figure 3a shows our initial theoretical model. The model has three latent variables: RCA, Working Memory (WM), and Inhibitory Control (IC), each measured through a number of manifest variables. The direction of the arrows indicates the direction of causal influence.

Model quality assessment takes three general steps. The first is to verify the relationship between manifest variables and the latent variables hypothesized to underlie them. This is done by first assessing unidimensionality, or the extent to which a change in the latent variable affects all manifest variables in the same direction. Unidimensionality is indexed by Dillon-Goldstein's rho (DG rho). A DG rho above 0.7 is favorable. Second, the relative contribution of the latent variable (vs. noise/task-specific factors) to each manifest variable is calculated. Indicators with a loading of less than 0.6 are not good indices of the latent construct.

Third, the cross-loadings of the manifest variables are checked to ensure that the loading of a manifest variable is indeed highest on the latent variable it is assumed to represent, and not on another latent variable in the model.

The first three steps are used to revise the model by dropping manifest variables or re-partitioning the latent variable constructs. The revised model is then re-checked.



a.
b.

Figure 3: Structure of initial path model (a) and revised model for production data only (b).

Once a viable model is obtained, the overall fit is assessed, which also includes the relationship between the latent variables. R^2 is reported for the latent variable of interest, and similar to simple regression models, indicates the amount of variance explained by the independent latent variables (WM and IC).

The part of the model that answers questions about the contribution of general cognitive functions on RCA is the inner model, or the links between latent variables. Its output is similar to that of any generalized linear model. Significance levels of path coefficients are estimated via bootstrapping with 1000 iterations.

Results

Model 1: General Model

Model 1 is based on the following theoretical assumptions: (a) the three WM tests measure a unified WM construct, (b) the three IC tests measure a unified inhibition construct, and (c) all the RCA scores measure a unified RCA construct.

Examination of this general model resulted in three main revisions: (a) Rhyme Probe was dropped because it did not contribute substantially to the latent WM construct (loading = 0.12) (b) The loadings and cross-loadings of the latent variables revealed that the IC variable had low unidimensionality (DG rho = .11). NoGo errors had a large negative loading on the IC latent variable (-0.93), while the other two IC manifest variables had positive loadings, indicating that the NoGo errors were measuring a fundamentally different construct. Thus, the IC latent variable was broken down into two variables: indirect competitive (CI; Semantic Blocking and Flanker) and global (GI; NoGo) to reflect two types of IC that have been shown to have different cognitive and neural underpinnings (Munakata et al., 2011). (c) While all measures of RCA in production had factor loadings above 0.6 on the RCA latent variable, RCA measures in comprehension had low factor loadings and low unidimensionality across Adj+ and Adj- conditions. This finding points to dissociation of the RCA construct in production and comprehension and calls for separate examination of the two tasks.

Model 2: Production

This model is shown in Figure 3b, and includes only manifest variables indexing RCA in production. RCA had a DG rho of 0.86, and high factor loadings for all four manifest variables. This finding supports defining RCA in language production as a unified construct, regardless of linguistic demands. DG rho's for WM and CI were 0.7 and 0.63 respectively, with R^2 of 0.51.

Table 1 shows the results of the bootstrapping on Model 2. The only latent variable that significantly predicted RCA in production was GI. Post-hoc modeling of Adj+ and Adj- manifest variables separately also revealed a reliable contribution of this variable to RCA in both trial types.

Table 1: Results of Bootstrapping for Model 2. CI = Competitive Inhibition; GI = Global Inhibition; WM = Working Memory.

Latent IV	Path Coeff.	Bootstrapping Means	SE	Lower 95% CI	Upper 95% CI
WM	0.18	0.20	0.20	-0.24	0.54
CI	0.26	0.18	0.28	-0.42	0.57
GI	-0.56	-0.54	0.18	-0.80	-0.21

Analysis of eye-tracking data in production showed that participants' average competitor advantage scores were positively correlated with RCA accuracy in both the Adj+ and Adj- condition (r = .45 and .58, respectively; p's < .05), and that eye-tracking performance itself was highly correlated across the two conditions (r = .96, p < .001).

Analysis of Comprehension Data

In the path model, the two measures of RCA in comprehension had opposite loadings on the RCA latent variable, thus forcing us to examine them separately. To understand why the eye-tracking measures in comprehension had opposite effects on the RCA variable, we first examined the correlation between participants' target advantage scores in the Adj+ and Adj− conditions. The two measures showed a significant negative correlation (r = -.47, p = .01), in contrast to the production results. Nearly half of the participants (13 of 28) showed a negative target advantage score in one condition and a positive score in the other, indicating that in one condition, participants systematically interpreted the critical cohort phoneme as the onset of the competitor, instead of the target. Recall that in the Adj+ condition, the target word was always an adjective and its cohort competitor was always a noun, while in the Adj− condition, the opposite was true. Thus, it appears that these participants adopted a strategy of always interpreting the cohort phoneme as being from the same part of speech.

Because our main question of interest was whether RCA in production and comprehension rely upon the same cognitive processes within an individual, we directly compared performance in these two domains within the subset of participants who were clearly engaging in RCA during comprehension (Adapters; n = 11). While no relationship was found between production and comprehension on Adj+ trials, on Adj− trials, these participants' eye-tracking performance in comprehension was significantly positively correlated with their production accuracy (r = .63, p < .05) and marginally positively correlated with their eye-tracking performance in production (r = .49, p = .12).

We also tested whether the same cognitive abilities were responsible for performance across conditions in Adapters. Due to the univariate nature of the dependent measures in these analyses, we used multilevel models, run separately on Adj+ and Adj− eye-tracking data, with all six cognitive tests included as predictor variables. While none of the cognitive measures predicted performance on Adj+ trials, RCA on Adj− trials was significantly predicted by performance on the Flanker (t = 6.2, p < .01) and NoGo tasks (t = 5.1, p < .05).

In summary, comprehension results differentiated two groups of individuals: those who flexibly adapted their processing to referential context and those who did not. In those who did, RCA abilities were correlated in production and comprehension, at least on Adj− trials, and in both cases they were well predicted by a measure of IC.

Discussion

This experiment tested three central questions: 1) Which cognitive processes underlie RCA? 2) Are these processes consistent across production and comprehension? 3) Do situations in which better RCA is marked by addition of an adjective differ from those in which it is marked by omission of an adjective? In answer to question 1, we observed a clear effect of inhibitory control across the domains of production and comprehension. In production, NoGo scores were predictive of better RCA for both Adj+ and Adj- trials. In comprehension, amongst Adapters, scores on both NoGo and Flanker tasks predicted performance on Adj− trials. These findings suggest a definite role for inhibitory control in RCA. Since common to both production and comprehension, this finding most likely points to the role of inhibitory control in capturing the critical contrast by preventing fixed attention to the target (hence ignoring the critical competitor).

Two lines of evidence can be used to answer question 2. First, the fact that inhibitory control played a role in RCA for both production and comprehension is evidence that the

RCA abilities are related in the two domains. Second, performance on Adj− trials was correlated between production and comprehension amongst Adapters. Together, these findings provide the first piece of evidence for a common basis of RCA in production and comprehension.

In answer to question 3, analyses on production strongly suggested that RCA ability was independent of trial type: Eye-tracking measures across Adj+ and Adj- trials were highly correlated, PLS-PM analysis revealed the RCA latent variable in production to be a highly coherent construct, and the same underlying cognitive mechanism, global inhibition, was implicated in both Adj+ and Adj− trials. In comprehension, more variability across conditions and across participants was observed. The PLS-PM model showed a lack of unidimensionality across comprehension conditions, and upon further inspection of the data, it was found that nearly half of the participants did not use referential context. Even within Adapters, only RCA on Adj− trials correlated with production and was predicted by measures of inhibitory control. This is in part due to low internal reliability of RCA in comprehension (ρ* = .34 for Adj− and ρ* = -.16 for Adj+)[1], as also reported by Ryskin et al. (2015), which stands in sharp contrast to the high split-half reliability of RCA in production (ρ* = .97 in both conditions). However, our results suggested that apart from consistency issues, listeners did often default to a fixed strategy, as opposed to flexibly changing their behavior according to context, as they did in production.

In summary, these results represent the first evidence for shared underlying cognitive mechanisms of pragmatic processing in production and comprehension. In production, this mechanism, global inhibition, was recruited regardless of the particulars of the referential context, while in comprehension, results were less uniform across conditions, pointing to specific strategies adopted by listeners in locating the referent. These findings provide insights into the cognitive processes that drive pragmatic use during spoken language comprehension and production, and help to situate pragmatic processing within a larger and more general cognitive framework.

Acknowledgments

This research was supported by a Ruth L. Kirschstein NRSA Postdoctoral Fellowship (F32DC015390) to AMT & the Therapeutic Cognitive Neuroscience Fund.

References

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433-436.

Brown-Schmidt, S. (2009). The role of executive function in perspective taking in online language comprehension. *Psychonomic Bulletin & Review*, *16*, 893-900.

Deutsch, W., & Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition*, *11*, 159-184.

Freedman, M. L., & Martin, R. C. (2001). Dissociable components of short-term memory and their relation to long-term learning. *Cognitive Neuropsychology*, *18*, 193-226.

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. Morgan (eds.) Syntax and Semantics Volume 3: Speech Acts. New York: Academic Press.

Kessels, R. P., Van Zandvoort, M. J., Postma, A., Kappelle, L. J., & De Haan, E. H. (2000). The Corsi block-tapping task: standardization and normative data. *Applied Neuropsychology*, *7*, 252-258.

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, *46*, 551-556.

Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O'Reilly, R. C. (2011). A unified framework for inhibitory control. *Trends in Cognitive Sciences*, *15*, 453-459.

Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology*, *58*, 220-249.

Nozari, N., Trueswell, J. C., & Thompson-Schill, S. L. (in press). The interplay of local attraction, context and domain-general cognitive control in activation and suppression of semantic distractors during sentence comprehension. *Psychonomic Bulletin & Review*.

Ryskin, R. A., Benjamin, A. S., Tullis, J., & Brown-Schmidt, S. (2015). Perspective-taking in comprehension, production, and memory: An individual differences approach. *Journal of Experimental Psychology: General*, 1*44*, 898-915.

Sanchez, G., Trinchera, L., & Russolillo, G. (2013). plspm: tools for partial least squares path modeling (PLS-PM). R package version 0.4, 1.

Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language*, *54*, 199-227.

Sedivy, J. C. (2005). Evaluating Explanations for Referential Context Effects: Evidence for Cricean Mechanisms in Online Language Interpretation. In Trueswell, J. C., & Tanenhaus, M. K. (eds). *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. MIT Press.

Wardlow, L. (2013). Individual differences in speakers' perspective taking: The roles of executive control and working memory. *Psychonomic Bulletin & Review*, *20*, 766-772.

[1] ρ* = Spearman-Brown adjusted correlation, a measure of split-half reliability