From visual prominence to event construal: influences (and non-influences) of eyegaze

Lilia Rissman (l.rissman@let.ru.nl)
Center for Language Studies, Erasmusplein 1
VR6541 Nijmegen, the Netherlands

Susan Goldin-Meadow (sgsg@uchicago.edu)
Department of Psychology, 5848 S. University Ave.
Chicago, IL 60637 USA

Amanda Woodward (woodward@uchicago.edu)
Department of Psychology, 5848 S. University Ave.
Chicago, IL 60637 USA

Abstract

Perceptual aspects of events, such as the visual prominence of event participants, have been shown to influence how people describe events. We investigate the relationship between such perceptual effects and patterns of eyegaze, focusing on a little-explored perceptual manipulation: the extent to which an event participant is wholly or partially visible. Using an eyetracking method, we found a correlation between this perceptual contrast and patterns of eyegaze at the beginning of the event but not the end. This finding supports the view that early visual attention to events has important downstream consequences for event conceptualization and linguistic description.

Keywords: events; construal; agency; eyetracking; visual perception; conceptualization

Introduction

When people describe events, they have a variety of linguistic options at their disposal: depending on your goals and interests, for example, you could tell your dinner guests *I burnt the soufflé* or simply *the soufflé got burnt*. Decades of research have addressed how a multitude of conceptual, linguistic and discourse factors influence language production choices (see Bock & Ferreira, 2014 for review). The last decade has seen increased attention to how visual perceptual aspects of event apprehension affect these choices (Coco & Keller, 2015; Griffin & Bock, 2000; Gleitman, January, Nappa & Trueswell, 2007; Myachykov, Garrod & Scheepers, 2012; Vogels, Krahmer & Maes, 2013). If, for example, a participant happens to look at one entity in an event before another, is that entity more likely to be mentioned as the Subject of the sentence? In the current study, we test the relationship between visual prominence in an event and how participants apprehend events through their eyegaze (e.g., when participants fixate on entities in an event and for how long). Previous work shows that for simple agentive events such as a person tipping over a book, participants are more likely to describe these events using passive voice when only the agent's hand is visible than when the face and torso are also visible (Rissman, Woodward & Goldin-Meadow, under revision). We asked in this study whether this linguistic effect is reflected in patterns of eyegaze. We measured participants' eye movements while viewing events that differed in terms of how much of the agent was visible, and found that eyegaze patterns distinguished these two types of events at the beginning of an event but not at the end. Our results provide support for the proposal that early fixations to individual scene elements play an important role in shaping language production (Gleitman et al., 2007; Myachykov et al., 2012; cf. Griffin & Bock, 2000; Bock, Irwin & Davidson, 2004), and are consistent with claims that early fixations guide how events are conceptualized (Bock & Ferreira 2014). At the same time, only one of our eyegaze measures correlated with degree of occlusion of the agent, informing our understanding of ways in which eyegaze does not track cognitive processing.

The influence of the eyes in language production

While a rich literature has addressed how conceptual features such as animacy affect linguistic descriptions (Bock, Loebell & Morey, 1992; Branigan, Pickering & Tanaka, 2008; among others), we understand less about the influence of visual perception on event conceptualization and linguistic encoding. In English, the Subject is usually also the topic (Lambrecht, 1994). Thus the active and passive sentences *the cat tipped over the water* and *the water was tipped over* share the same relational structure (i.e., who did what to whom) but differ in terms of what the sentence is primarily "about." Influential studies by Bock and colleagues have shown no relationship between early patterns of eyegaze and Subject selection (Griffin & Bock, 2000; Bock, Irwin, Davidson & Levelt, 2003; Bock et al., 2004). Griffin & Bock (2000), for example, tracked participants' eyegaze while viewing static illustrations of two animate entities in an event. They found that entities that were fixated first were not more likely to be mentioned as the Subject of a sentence.

Subsequent studies, however, have shown relationships between early eyegaze and language production. In Gleitman et al. (2007), participants viewed illustrations with two animate entities, e.g. a cat licking a dog. Just prior to viewing the illustration, a square flashed briefly over the space where one of the entities would ultimately appear. Overall, speakers usually mentioned the agent as the Subject, but when the square had flashed over the patient, participants were more likely to produce a passive description (e.g., *the dog was licked by the cat*) than when the square had flashed over the agent. Directing visual attention to the patient thus increases the likelihood of that entity surfacing as Subject. Similar effects have been reported by Forrest (1996), Tomlin (1997), Myachykov et al. (2012), Vogels et al. (2013) and Coco & Keller (2015), among others.

Such visual perceptual effects have been primarily explained in terms of two mechanisms, which are not mutually exclusive: lexical access and conceptual Figure-Ground assignment. In the first mechanism, when a participant views a scene of a dog and a cat, looking at the dog first makes the word *dog* more active and subsequently more likely to be mentioned as Subject. In the second mechanism, participants are more likely to construe an event as being "about" a specific entity if this entity is fixated first -- in other words, that entity is the conceptual Figure.

This literature raises several issues: first, the fact that not all studies have found effects of early visual perception on language calls for exploration into a wider range of visual contrasts, to better understand the robustness of such effects. Second, if visual prominence affects conceptual Figure-Ground assignment, we would expect visual effects to be found even for non-linguistic tasks. Third, what is the relationship between visual prominence and the eyegaze itself: at what stages of event apprehension (i.e., beginning, middle, end) does eyegaze most reflect conceptual and linguistic processing? We address here each of these issues.

Research approach

We extend previous research showing that English-speaking adults produce more passive descriptions when less of the agent is physically visible (Rissman et al., under revision). These authors found in four separate studies that when participants describe videos of agentive events such as a person knocking over a book, they are more likely to produce a passive description such as *the book was tipped over* when the frame of the video contains only the hand of the agent, and not the face and torso as well (see Figure 1 for examples of "Full" visibility vs. "Partial" visibility events). Animacy and agency are both associated with higher conceptual accessibility (Bock & Warren, 1985). Nonetheless, this finding shows that perceptual information can override *both* the agency and animacy cues in determining which referrent is mentioned as Subject. We are not aware of previous studies that have manipulated degree of occlusion of event referents.

Rissman et al. conducted a follow-up comprehension task which indicated that obscuring the face/torso leads participants to conceptualize the event as being more "about" the patient, i.e. a construal where the patient is the Figure. In this task, participants were shown a Full visibility event alongside a Partial visibility event such as in Figure 1. Participants were given either an active sentence or a passive sentence and were told to select which video best matched the sentence. In two separate studies, participants were more likely to select the Full video given an active sentence rather than a passive sentence. This provides further evidence that when the agent's face is obscured, participants construe the event as being more "about" the patient. In an additional rating study, participants judged the agent to have the same degree of animacy in the Full and Partial videos.

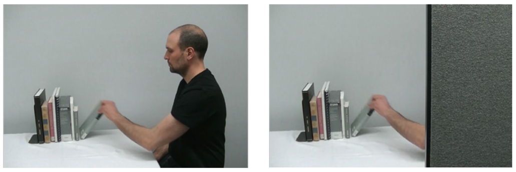

Figure 1: Example videos from the Full-Agent and Partial-Agent conditions. In this scene type, the person tips over the book onto the table.

If Full and Partial events lead to alternate event conceptualizations, this difference should be present even when participants are doing a non-linguistic task. We asked, therefore, whether patterns of eyegaze would be different for Full and Partial events, in the absence of a language production task. Differing event construals need not be reflected in patterns of eyegaze. These construals may instead be the result of higher-level conceptual-pragmatic reasoning with no correlate at the mechanistic level of the gaze. Participants may reason, for example, that if the experimenters did not choose to show the entire agent in the Partial events, then the agent must be relatively unimportant, or there may be a general dispreference for small subparts of an entity to be conceptualized as the Figure.

We also investigated *when* eyegaze would differ over the course of viewing Full vs. Partial events (if at all). We hypothesized two possible ways in which patterns of looking might differ: first, that participants would direct their early looks to the patient more often for Partial events than for Full events, consistent with previous findings on the role of early fixations. In Partial events, there are fewer entities to look at, since the highly salient face is absent. This does not automatically mean, however, that participants will be faster to look at the patient for Partial than for Full events. As a marker of animacy, a hand is still more conceptually prominent than an inanimate object, and it may be that participants in the Partial condition direct their early looks to the hand at the same rate as participants in the Full condition direct their early looks to the face/body/hand.

Our second hypothesis about the timing of any eyegaze effects was that participants would shift their attention in different ways for Full vs. Partial events. Pilot studies showed that participants fixate longer on the patient and hand at the end of the event than at the beginning: thus participants shift their attention more toward the patient and the hand, and away from other aspects of the scene, at the end than at the beginning of the event. Since Partial events are construed more from the patient's perspective, we predict that participants viewing Partial events will shift a higher proportion of their looking towards the patient than participants viewing Full events.

Method

Participants

32 native English-speaking adults from the University of Chicago community participated (M = 21, F = 11; mean age = 21). An additional 11 participants were tested but were excluded due to being non-native English speakers (N = 5), or due to technical problems with the eyetracker (N = 3), or because the eyetracker failed to record eyegaze data for over 40% of the samples for that participant (N = 3). All participants received either $10 or course credit for their participation.

Design, Materials & Procedure

We compared patterns of fixation for agentive events where the person's face and torso were visible (Full-Agent condition) against events where only the person's hand and forearm were visible (Partial-Agent condition). Examples of each type of video are shown in Figure 1.

There were eight types of scenes in each of these two conditions, e.g. a person tipping over a book, a person pushing a ball into a cup, and a person opening the lid of a box. In each video, a person sat with their hand resting on a table. They performed the action and returned their hand to the initial resting position. Videos were edited so that: 1) the initial and final resting periods lasted exactly 1 sec each, and 2) the motion component of the event (the person moving their hand to the object, acting on the object and returning their hand to the table) lasted exactly the same amount of time for the Full-Agent and Partial-Agent videos (3-4 seconds for each type of scene). Partial-Agent videos were filmed with the person sitting behind a screen. In Full-Agent videos, only the profile of the person's face was visible.

In addition to these Agent videos, participants also viewed No Agent videos, in which a person sat motionless at a table while the object underwent a change on its own. These No Agent videos featured the same eight types of scenes as the Agent videos (e.g., a book tipping over, a ball rolling into a cup, the lid of a box falling open). As with the Agent videos, there were variants where the person's face and torso were visible, as well as where only the forearm was visible. The purpose of these stimuli was to prohibit participants from expecting that the person would move, and thereby make anticipatory looks. Videos were edited so that the duration of the motion of the object (e.g., the book moving from vertical to horizontal position) was the same for No Agent videos as for Agent videos. For each type of scene, No Agent and Agent videos were also the same overall length.

Participants viewed 32 videos in total, or four videos for each of the eight types of scenes. We constructed four random orders of these videos, and participants were evenly distributed across these four orders. A centrally-located fixation cross was shown for 1.5 sec between each video.

Videos were shown on a Tobii T60 XL eyetracker sampling at 60 Hz. Participants were asked to watch the videos as if they were watching TV, and to look at the fixation cross between each video.

Data analysis

We report here eyegaze data for the Agent videos. We analyzed fixations to three areas of interest (AOIs): the Body, the Hand and the Patient. Examples of each of these AOIs are shown in Figure 2 for both Full and Partial videos. The Patient AOI included the regions that the object was occupying at the beginning and at the end of the event.

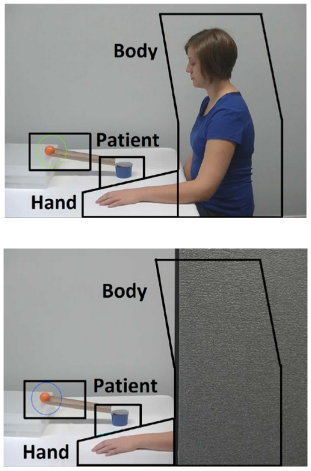

Figure 2: Example AOIs from the Full-Agent and Partial-Agent conditions. In this scene, the person pushes a ball down a ramp into a cup

We analyzed fixations to these AOIs for only the beginning and end of the event, when neither the person nor the object was in motion. Adults need about 200 ms to plan a saccade (Duchowski, 2007). We therefore analyzed an 800 ms segment at the beginning of the event (200 ms - 1000 ms; Beginning segment) as well as the final 800 ms of the event (End segment).

Eyegaze data were filtered using the Tobii I-VT filter. Individual trials were excluded if there was less than 67% gazepoint data for that trial; 4% of trials were excluded for this reason. For those remaining trials, gazepoint data was available for 94% of the samples in the trial.

We analyzed two dependent measures: 1) proportion of participants looking to each of the AOIs early in the trial and 2) total fixation duration to each of the AOIs during the Beginning and End segments. Fixation duration was calculated as a proportion of the percentage of the trial for which there was gazepoint data available.

Results

Early looking

Our first prediction was that participants in the Partial condition would be more likely to direct their early attention

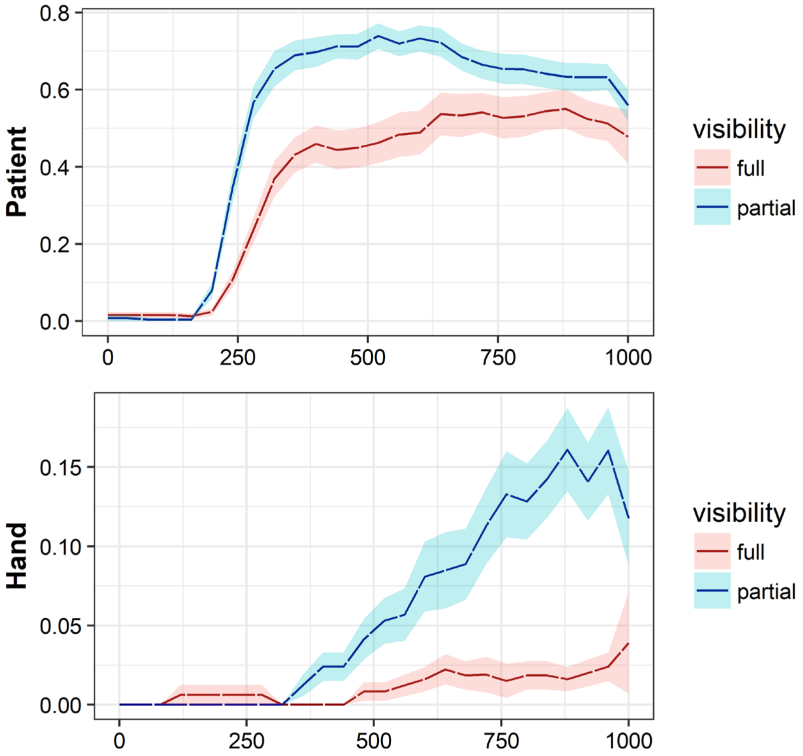

Figure 3: Average proportion of participants looking to the Patient and Hand AOIs in the Full and Partial visibility conditions. The timepoint 0 marks the beginning of the trial. Note different y-axis scales for Patient vs. Hand. Shaded regions show 95% confidence intervals.

to the Patient than participants in the Full condition. Figure 3 shows the proportion of participants who were fixating on the Patient and Hand AOIs in the Full and Partial conditions in the first 1000 ms of each trial, binned in 40 ms increments.

Figure 3 shows a steep upward incline in looking to the Patient AOI in both the Full and Partial conditions, but an even steeper incline for the Partial condition. By contrast, looking to the Hand AOI increases slowly over the course of the beginning of the trial, with looking to the Hand ultimately more frequent in the Partial than in the Full condition.

Adults need about 200 ms to plan a saccade, and saccades themselves last 10-100 ms (Duchowski, 2007). To gain insight into participants' earliest fixations, we measured the proportion of participants who were looking at the Patient and the Hand AOIs in the time window 280-320 ms after the start of the trial. These values are shown in Table 1.

We used mixed model logistic regression and the *lme4* package for R (Bates & Maechler, 2009) to model the likelihood of participants looking to the Patient during this time window. The best-fitting model included participant and item random intercepts and the Visibility fixed effect (Full vs. Partial). Participants in the Partial condition were significantly more likely than participants in the Full condition to be fixating on the Patient during this early time window (β = 1.71, SE = .22, p < .001). The absence of the face and torso of the Agent does not lead participants in the Partial condition to redirect their attention to the Hand early in the trial: indeed paticipants looked to the Hand almost not at all in the first 400 ms of the trial, in either condition. Our prediction was thus confirmed that participants in the Partial condition are more likely to direct their early looks to the Patient than participants in the Full condition. These results indicate that differences in construal do in fact correlate with differences in eyegaze, and that these differences emerge early in the trial.

Table 1: Proportion of participants fixating on each AOI in each condition during the 280-320 ms time window

AOI	Visibility	prop Ss (95 CI)
Patient	Full	.24 (.06)
	Partial	.57 (.05)
Hand	Full	.004 (.008)
	Partial	0 (0)
Body	Full	.33 (.06)
	Partial	.04 (.02)

Fixation duration

Figure 4 shows proportional fixation duration at the beginning vs. the end of the video for each of the AOIs, in the Full and Partial visibility conditions. Consistent with a

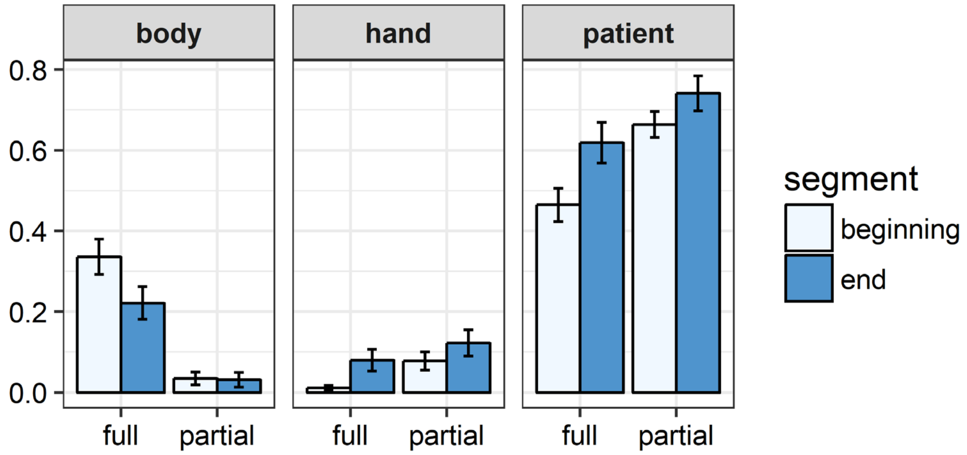

Figure 4: Proportional fixation duration to the Body, Patient and Hand AOIs in the beginning and end segments of the video in the Full and Partial visibility conditions. Error bars show 95% confidence intervals.

previous pilot, participants fixated on the Patient and the Hand AOIs more at the end of the video than at the beginning, in both the Full condition (beginning = .47 vs. end = .70) and the Partial condition (beginning = .74 vs. end = .86). Our research question was whether participants would shift their attention toward the Patient and Hand in different ways in the Full vs. Partial condition: if the Partial videos are more "about" the Patient, then participants might shift a greater proportion of their gaze toward the Patient.

To address this question, we calculated for each trial the difference between the fixation duration at the end vs. the beginning of the trial for each of the Patient and Hand AOIs. We then divided this difference measure by the proportion of the beginning segment in which participants were *not* fixating to the Patient or Hand, for each of the Full and Partial conditions (for Full: 1-.47 = .53; for Partial: 1-.74 = .26). This "proportional shift" statistic is shown in Figure 5.

Impressionistically, there appears to be an effect of AOI, with participants shifting relatively more of their gaze to the Patient than to the Hand, but no effect of Visibility: participants appeared to shift their gaze in the same way in the Full and Partial conditions. We analyzed these data using linear regression, which confirmed these impressions: there was a significant effect of AOI ($\beta = .14$, SE = .07, $p = .05$) but no effect of Visibility ($\chi^2(1) = .06$, $p > .1$) or interaction between AOI and Visibility ($\chi^2(2) = .06$, $p > .1$). Given the null effect of Visibility, we do not find support for our hypothesis that participants shift their attention differently in the Full vs. Partial conditions. Thus we did not find that differing construals were correlated with different patterns of eyegaze at the end of the events.

Discussion

We investigated the relationship between visual prominence, here the degree to which an agent's body is visible, and participants' visual apprehension of events, as measured through eyetracking. We compared patterns of eyegaze for highly Agent-oriented events (Full-Agent videos) against eyegaze for more Patient-oriented events (Partial-Agent videos), as distinguished by previous linguistic findings (Rissman et al., under revision). We found that a greater proportion of participants directed their early looks to the Patient in the Partial condition than in the Full condition, with participants in either condition looking almost not at all to the Hand at the beginning of the event. By contrast, we found no difference between the Partial and Full condition in terms of how participants shifted their attention from the beginning of the events to the end.

The finding that early fixation distinguishes these events is consistent with Gleitman et al. (2007) and Myachykov et al. (2012), who found that cuing participants' initial attention influenced their subsequent linguistic descriptions. Our results are also consistent with the proposal that early fixation influences event conceptualization, specifically Figure-Ground assignment (see Bock & Ferreira 2014; Vogels et al. 2013). Our study cannot, however, determine the direction of causation between eyegaze on the one hand and event

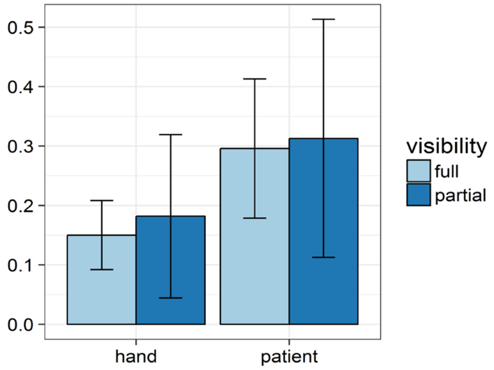

Figure 5: Proportion of attentional shift to the Patient and Hand AOIs in the Full and Partial visibility conditions. Error bars show 95% confidence intervals.

conceptualization on the other: it may be that participants in the Partial condition directed their early looks to the Patient, which led to patient-oriented construals. Alternatively, participants may have encoded a high-level construal for each of the Full and Partial videos within milliseconds of viewing them, which led to the patterns of eyegaze we observed. Further experimentation is needed to distinguish these explanations.

Another step for future research is to test whether early fixations to the Patient are correlated with more passive descriptions, where the patient is the Subject. Rissman et al. (under revision) found that participants produced both active and passive descriptions of Partial videos. If fixating initially on the Patient leads to more patient-oriented construals, we predict that it should also be correlated with more passive descriptions.

Participants in our study were not doing a linguistic task, yet we see correlations with early fixation parallel to those reported for linguistic tasks. One explanation for this similarity is that early fixation helps determine the Figure, and if participants are doing a linguistic task, the Figure is more likely to be mentioned as the Subject. This interpretation comports with evidence that initial apprehension of a static image influences Figure-Ground assignment, e.g. which individual is perceived in the classic ambiguous young lady/old woman drawings (Georgiades & Harris, 1997; Vecera, Flevaris & Filapek, 2004). Understanding the relationship between eyegaze and event conceptualization is important not only for building theories of grammatical encoding during event processing, but also for understanding representations of Figure-Ground assignments in populations that cannot provide linguistic descriptions, such as infants.

In the literature on grammatical encoding, the formulation of the "message" component is relatively under-studied. When we describe events, formulating a message involves visual perceptual processing, which then serves as input to a conceptual representation of the event. This study contributes to our understanding of how visual apprehension can have important consequences for higher-level event concepts.

Acknowledgments

This research was supported by a postdoctoral supplement awarded to Lilia Rissman and Susan Goldin-Meadow as part of NIH RO1 DC000491 (PI: Goldin-Meadow). Thank you to Chiara Pandolfi for her hard work as a research assistant, and all research participants for their time and attention.

References

Bates, D., & Maechler, M. (2009). lme4: linear mixed effects models using S4 classes.

Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition, 21*(1), 47-67.

Bock, K., & Ferreira, V. S. (2014). Syntactically speaking. In V. S. Ferreira, M. Goldrick, & M. Miozzo (Eds.), *The Oxford handbook of language production*. Oxford, England: Oxford University Press.

Bock, K., Irwin, D. E., & Davidson, D. J. (2004). Putting first things first. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: eye movements and visual world*. New York: Psychology Press.

Bock, K., Irwin, D. E., Davidson, D. J., & Levelt, W. J. M. (2003). Minding the clock. *Journal of Memory and Language, 48*(4), 653-685.

Bock, K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: bridging the syntactic cleft. *Psychological review, 99*(1), 150-171.

Branigan, H. P., Pickering, M. J., & Tanaka, M. (2008). Contributions of animacy to grammatical function assignment and word order during production. *Lingua, 118*(2), 172-189.

Coco, M. I., & Keller, F. (2015). Integrating mechanisms of visual guidance in naturalistic language production. *Cognitive Processing, 16*(2), 131-150.

Duchowski, A. T. (2007). *Eye tracking methodology: Theory and practice*. London: Springer.

Forrest, L. B. (1996). Discourse goals and attentional processes in sentence production: the dynamic construal of events. In A. Goldberg (Ed.), *Conceptual structure, discourse and language* Stanford, CA: CSLI Publications.

Georgiades, M. S., & Harris, J. P. (1997). Biasing Effects in Ambiguous Figures: Removal or Fixation of Critical Features Can Affect Perception. *Visual Cognition, 4*(4), 383-408.

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language, 57*(4), 544-569.

Griffin, Z. M., & Bock, K. (2000). What the Eyes Say About Speaking. *Psychological Science, 11*(4), 274-279.

Lambrecht, K. (1994). *Information structure and sentence form: topic, focus, and the mental representations of discourse referents*. Cambridge England; New York: Cambridge University Press.

Myachykov, A., Garrod, S., & Scheepers, C. (2012). Determinants of structural choice in visually situated sentence production. *Acta Psychologica, 141*(3), 304-315.

Rissman, Woodward & Goldin-Meadow (under revision). Occluding the face diminishes the conceptual accessibility of an animate agent.

Tomlin, R. S. (1997). Mapping conceptual representations into linguistic representations: The role of attention in grammar. In J. Nuyts & E. Pederson (Eds.), *Language and conceptualization* (pp. 162-189). Cambridge; New York, NY: Cambridge University Press.

Vecera, S. P., Flevaris, A. V., & Filapek, J. C. (2004). Exogenous spatial attention influences figure-ground assignment. *Psychological Science, 15*(1), 20-26.

Vogels, J., Krahmer, E., & Maes, A. (2013). Who is where referred to how, and why? The influence of visual saliency on referent accessibility in spoken language production. *Language and Cognitive Processes, 28*(9), 1323-1349.