# Grammatical aspect in language production: Using gesture to reveal event representations

**Fey Parrill (fey.parrill@case.edu)**
Department of Cognitive Science, 10900 Euclid Ave.
Cleveland, OH 44106 USA

**Benjamin Bergen (bkbergen@cogsci.ucsd.edu)**
Department of Cognitive Science, 9500 Gilman Dr.
La Jolla, CA 92093 USA

**Patricia Lichtenstein (plichtenstein@ucmerced.edu)**
Department of Cognitive and Information Science, 5200 North Lake Rd
Merced, CA 95343 USA

**Abstract**

Grammatical aspect is a pervasive linguistic device that, according to linguistic analyses, allows speakers to encode different ways of construing events. For instance, the progressive (*I am writing a book*) is thought to reflect increased focus on the internal details of an event, as contrasted with the perfect (*I have written a book*). However, there is to date no experimental evidence that speakers in fact think about the same events differently when they describe them using progressive versus perfect aspect. We used co-speech gesture as a means to investigate what speakers' event representations are like when they produce perfect versus progressive utterances. We found that progressive event descriptions were accompanied by longer-lasting and more complex gestures than perfect event descriptions, but only when participants described events originally presented in the progressive. This evidence suggests that people are actually construing events differently when they use different grammatical aspects.

**Keywords:** Gesture, aspect, language production, mental simulation.

## Introduction

Most of the world's languages mark the structure of events, using a linguistic device known as *grammatical aspect*. A variety of possible aspectual distinctions are attested, but among the more common ones is to mark whether an event is to be understood as ongoing or completed. For instance, the English *progressive* aspect, as in *I was writing a book* indicates that the event is to be conceived of as ongoing, while the *perfect*, seen in I *had written a book*, marks the event as completed. (Aspect is distinct from tense, which marks when an event occurred relative to the time an utterance is produced.) Linguists who have looked closely at grammatical aspect have argued that using the progressive, as contrasted with the perfect, increases focus on the internal structure of the event, or increases the granularity or detail with which people think of events (e.g., Comrie, 1976; Dowty, 1977; Langacker, 1991).

This intuition jibes well with native speaker intuition, but to date there has been no experimental work on the question. When speakers use different aspectual forms, are they actually thinking about events differently? This is the basic question we address in the research below.

While there hasn't been work on how speakers represent events while producing different grammatical aspects, there has been a bit of recent work on language comprehension. Several studies (Madden & Zwaan, 2003; Bergen & Wheeler, 2010) have found evidence that aspect modulates how comprehenders represent described events, in ways quite compatible with the predictions that the linguistics literature on aspect makes. For instance, Bergen & Wheeler (2010) compared the extent to which comprehenders represent the details of motor actions when presented with progressive sentences (*John is opening the drawer*) and perfect sentences (*John has opened the drawer*). They found an increased action-sentence compatibility effect (Glenberg & Kaschak, 2002) with progressive sentences; progressive sentences led to faster performance of subsequent compatible actions, but perfect sentences did not show such an effect. The authors interpreted this as evidence that the progressive encourages comprehenders to mentally simulate the nucleus of an event (in this case, a motor action) with greater detail than the perfect does.

The idea that aspect modulates mental simulation in comprehension is based on the more basic hypothesis that understanding language involves generating perceptual and motor representations of what is being talked about. Substantial theoretical and empirical work now supports this view (Barsalou, 2008, 2009; Bergen & Chang, 2005; Bergen, Lindsay, Matlock, & Narayanan, 2007; Bergen & Wheeler, 2010; Glenberg & Kaschak, 2002; Hostetter & Alibali, 2008, 2010; Kaschak & Glenberg, 2000; Kaschak et al., 2005; Matlock, 2004; Zwaan, 1999; Zwaan, Stanfield, & Yaxley, 2002). As a result, we adopt its terminology for the purposes of our exposition here. However, it is not essential to the study we describe below; when we talk about mental simulation of events, we could just as well be talking about mental representations of events that are not modality specific, as long as the representational format affords the possibility of different construals of events.

As mentioned above, while there is now some evidence that reading or hearing progressive sentences encourages a language user to simulate the internal details of an event, there has been no work to date on aspect and the mental operations of people *producing* language. It is possible that language production is like language comprehension in that speakers are more likely to produce progressive aspect when they are mentally simulating the event they are describing in greater detail; simulation-based theories usually assume that similar principles operate during comprehension and production (e.g., Barsalou, 2009, Bergen & Chang, 2005).

How can we assess whether people are thinking differently about events when they produce sentences using progressive versus perfect aspect? One valuable source of data is the gestures that speakers spontaneously produce along with speech. Such co-speech gestures are very closely linked, both in meaning and time, to the speech they accompany (McNeill, 1992, 2005). As a consequence, they provide a window—one distinct from the speech stream itself—onto the mental representations that underlie language. Co-speech gestures have long been thought to reflect imagistic aspects of thinking (e.g., Kita & Özyürek, 2003), but recent work has begun to link research on gesture to simulation-based theories of language. Hostetter and Alibali (2008, 2010) have suggested that gestures might reflect simulations carried out during language production. That is, perceptual and motor representations that automatically become active during language production might be the source of these gestures.

If gestures reflect underlying processes of simulation, then it stands to reason that gesture may shed light on whether aspectual choices reflect simulation differences during language production. In particular, gestures that accompany progressive-marked speech might reflect greater intensity or detail of focus on events that the speaker is thinking about. This greater attention could result in gestures that last longer and are more likely to encode information about the internal structure of events. Some corroboration for this possibility comes from work by Duncan (2002), who found that when the progressive appeared in narrative speech, co-occurring gestures were longer and more complex.

Gestural complexity can come in several forms, but a frequent characteristic is some kind of iteration or repetition. For example, if a narrator is describing someone walking, her co-occurring gesture might encode only the trajectory of motion (a straight line), or the gesture might also encode the internal structure of the event, by (for example) having the fingers wiggle while the hand moves in a straight line. The second gesture would be considered more complex, as it encodes an additional semantic feature (in this case, manner of motion: see Duncan, 2002; and see Kita & Özyürek, 2003 for more on manner of motion). In short, increased complexity, in the form of repeated or iterated gestures, can indicate a focus on the internal structure of an event.

The current study uses gesture in language production to assess the extent to which produced aspects reflect different construals of the described event. In doing so, it makes two main contributions. First, while there has been some work on how aspect affects event construal in comprehenders, this is the first work to address the reverse relation, in speakers. Second, we replicate and extend Duncan's (2002) gesture results by asking whether the same patterns can be observed for a carefully controlled set of events, where the same events are presented with different aspects.

## Study

If speakers use progressive aspect when they are focusing on the internal structure of events, gestures that accompany utterances marked with the progressive may reflect that fact. Specifically, they make take longer to produce and they may be more likely to encode semantic aspects of the event-internal structure.

To assess this possibility, we presented participants with a series of texts in which half of the events appeared in the past progressive, and half in the past perfect. We then asked participants to re-describe these texts to a listener, a task that tends to elicit gesture. This design allows us to look at two things. First, we can assess the relationship between the aspect of presentation (that is, the grammatical aspect in which an event appeared in the stimulus) and a participant's verbal and gestural behavior. Second, we can assess the relationship between the aspect of production (that is, the grammatical aspect a participant chooses to use when describing an event) and a participant's gestural behavior.

If, as other work suggests, mental simulation occurs during language comprehension, then reading sentences that vary in aspect should create varying representations during encoding. Those representations then form the content used in utterances subsequently formulated during language production. So presentation aspect might affect gestures produced during event retelling. Moreover, if production involves simulation as well, and if the aspect produced reflects differences in event construal, then production aspect might be affected by presentation aspect, and even when it does not, progressive aspect in production might correlate with more complex and longer lasting gestures.

### Method

Participants read a series of short texts, then described them to a partner.

**Materials** We created a set of fifteen written stories. Each consisted of seven sentences, the first of which was a scene-setting sentence that always appeared in the simple past. The remaining sentences each described an event likely to evoke gesture (a motion event). These events appeared in either the past progressive (e.g., *was floating*) or the past perfect (e.g., *had floated*). Aspect in each story alternated with each sentence: if the first sentence was past progressive, the next would be past perfect. We created two versions of each story (A and B), with aspect reversed for each target event. That is, in the A version, the first target event appeared in the past progressive, while in the B

version that same event appeared in the past perfect. We chose to alternate aspect within each story rather than across stories because stories containing only past perfect seemed awkward (they described a series of things that event participants *had done*). We used past progressive and past perfect (unlike the present progressive and present perfect found in Bergen & Wheeler (2010), for instance) because the past tense made the narratives seem more like typical stories. Two versions of each story were created so that we would have data for each event in both aspects. A sample text is found below. Target event verbs are in bold.

> There was a trail through the woods. A woman **was hiking/had hiked** down a little hill on the winding path. Afterwards, she **had walked/was walking** up to a little stream. A big stick **was floating/had floated** down the stream in front of a pair of ducks. The ducks **had paddled/were paddling** around it. Then they **were bobbing/had bobbed** around a bend in the stream out of sight. The woman **had leapt/was leaping** along some rocks to the other side of the stream.

Each story contained the same number of human, non-human-animate, and inanimate entities. Stories were normed to ensure that each was comprehensible and that there were no differences in comprehensibility across A and B versions. Twenty participants rated the A version of the stories on a scale from 1 (very hard to understand) to 5 (very easy to understand), and a separate group of twenty participants rated the B versions on the same scale. We selected the ten stories with the highest mean ratings for use in the study (Version A mean rating = 4.08, SD .16, Version B mean rating = 4.03, SD .13). There was no significant difference in the mean ratings across the A and B versions of these ten stories: $t(20) = .86$; p (two-tailed) = .40.

Thirty-five Case Western Reserve University students (17 women) participated in the study for payment. All were native speakers of English. Following informed consent, participants were seated at a computer and read a story sentence by sentence, pressing a computer key to advance to the next sentence. After the final sentence, the participant was prompted to move over to the recording area and to describe the story to his or her partner. The partner came to the study with the participant and served as a listener. Participants were told to describe the story in as much detail as possible, and that the partner would take a comprehension quiz at the end of the study. Participants randomly received either the A or B version of each story. Three participants were dropped from the study because they never gestured, for a total of 32 participants (17 who received version A, 15 who received version B).

**Coding** One coder transcribed all utterances that matched one of the target events from the stimuli, using audio alone. The grammatical aspect of these utterances was then coded as either progressive or non-progressive (typically simple past). One coder carried out this analysis for the entire dataset, and a second coded 42% of the dataset. For categorical data in this study, agreement was calculated using Cohen's kappa. Landis and Koch (1977) suggest that a kappa value kappa above .61 indicates substantial inter-rater agreement. For utterance aspect, kappa was .86. Any gesture that accompanied an utterance describing a target event was then coded according to the following categories (after McNeill, 1992, 2005): *concrete iconic* (the shape or motion of the hands maps onto some aspect of the scene being described), *metaphoric iconic* (the shape or motion of the hands maps onto an abstract property, such as discourse content), *deictic* (a pointing gesture), *beat* (a rhythmic gesture with no obvious semantic content), or *self-adaptor* (a self touching gesture, such as scratching the nose). Given our research questions, only concrete iconic gestures (66% of the total gestures produced in this dataset) were analyzed further. These gestures were coded for the presence of iteration. As noted above, iteration is a repeated action of the hands and is frequently associated with manner of motion. One coder carried out gesture type and iteration coding for the entire dataset. A second coder coded 46% of the dataset. Agreement was good: $\kappa = .70$ for gesture type and $\kappa = .76$ for iteration. Finally, *gesture stroke duration* was coded for concrete iconic gestures. Gestures are typically made up of multiple phases (McNeill, 1992): a preparation phase, a stroke phase (defined as the effortful, meaningful portion of the gesture: McNeill, 1992), and a retraction phase. In addition to these phases, gestures often involve holds, and these holds can extend a gesture's overall duration. However, holds can arise because of a need to keep gesture and speech temporally synchronous, and therefore are not necessarily semantically meaningful. For these reasons, gesture stroke duration was used for this study (as it was for Duncan, 2002) rather than overall gesture duration. Using software that displays frames per second, gesture stroke onset and offset can be identified within a $30^{th}$ of a second, with good reliability. Gesture stroke duration was coded using Final Cut Pro: stroke onset and offset were recorded and then converted to a total duration value in milliseconds. One coder carried out this analysis for the full dataset. A second coder coded 25% of the dataset. The correlation between the two coders was .86, indicating very good agreement.

In summary, the final dataset contained all utterances that matched a stimulus target event, coded for verbal aspect. When an utterance was accompanied by a concrete iconic gesture, that gesture was coded for the presence of iteration, and the duration of that gesture's stroke phase was calculated.

## Results

Results are presented below in terms of presentation aspect (the aspect that stimulus sentences appeared in) and production aspect (the aspect of participants' utterances). Before turning to gesture, we analyzed the relationship between these factors. First, we asked whether presentation

aspect affected production aspect in general, regardless of whether a gesture was produced. To account for variation across participants, we examined the proportion of matches (utterances that agreed in presentation and production aspect) for each participant. Table 1 shows these proportions (SDs in parentheses).

Table 1: Mean proportion of matches by presentation and production aspect (all utterances).

| | **Production** | |
|---|---|---|
| **Presentation** | Perfect | Progressive |
| Perfect | 0.53 (.15) | 0.47 (.15) |
| Progressive | 0.39 (.08) | 0.61 (.06) |

Participants did not always perfectly reproduce what they read, which is not surprising given the design: participants read sentences alternating between perfect and progressive. A Wilcoxon Signed Ranks test[1] showed participants were not significantly more likely to produce perfect when presentation aspect was perfect: Z =.97, p (two-tailed) = .32. However, when presentation aspect was progressive, participants were significantly more likely to produce progressive: Z =3.41, p (two-tailed) = .0006. We next examined the proportion of matches for each participant when concrete iconic gestures were produced. These means were identical to those in Table 1.

Again, a Wilcoxon Signed Ranks test showed participants were not significantly more likely to produce perfect when presentation aspect was perfect: Z =1.16, p (two-tailed) = .25. However, when presentation aspect was progressive, participants were significantly more likely to produce progressive: Z =3.32, p (two-tailed) = .0009. This greater tendency to produce progressive after reading progressive may reflect the fact that progressive is more common in speech.

Finally, we examined the relationship between presentation aspect and frequency of concrete iconic gesture production, to determine whether stimulus aspect had any impact on participants' tendency to produce such gestures. We compared the mean proportion of target event descriptions accompanied by gesture as a function of stimulus aspect: a Wilcoxon Signed Ranks showed no significant difference between these proportions (Z =.35, p (two-tailed) = .73).

Our second set of analyses focused on the relationship between presentation aspect, production aspect, and gesture stroke duration. Figure 1 shows the mean gesture stroke duration for utterances according to presentation and production aspect.

[1] Presentation aspect is an independent variable but production aspect is a dependent variable, thus a two-factor ANOVA is not appropriate for these analyses (assumptions of independence are violated). For this reason, we report results of Wilcoxon Signed Ranks tests for non-parametric data (frequencies), and paired t-tests for durations.

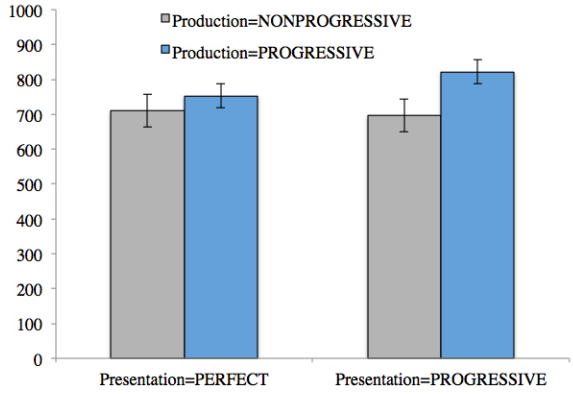


Figure 1. Mean gesture stroke duration (in msec) according to presentation and production aspect. Error bars show standard error.

Overall, gesture stroke duration was longer when the aspect of production was progressive: t(31) =2.04; p (two-tailed) = .049. To explore this result further, we first looked at the cases where presentation aspect was perfect, and compared mean gesture duration when production aspect was progressive and non-progressive. This test showed no significant difference between these means (t(31) =.87; p (two-tailed) = .38). When an event was originally presented using the perfect, people's gestures were not significantly longer when they themselves were recounting that event using progressive aspect than when they used perfect aspect.

Second, we turned to gesture duration when presentation aspect was progressive, again comparing progressive and non-progressive production aspect. This test showed a marginally significant difference between these means (t(31) =2.00; p (two-tailed) = .054). When an event had been originally described using progressive aspect, people made longer-lasting gestures when they themselves were recounting that event using progressive aspect than when they used perfect aspect. We also compared gesture stroke duration across production categories. There was no significant difference in duration as a function of presentation aspect: t(31) =1.14; p (two-tailed) = .26 for progressive utterances, and t(31) =.29; p (two-tailed) = .77 for non-progressive utterances.

We then turned to our second dependent measure, gesture iteration. Figure 2 shows the mean proportion of utterances accompanied by an iterated gesture, by presentation and production aspect. Overall, participants produced a larger proportion of gestures with iteration when aspect of production was progressive: Z = 2.32, p (two-tailed) = .02.

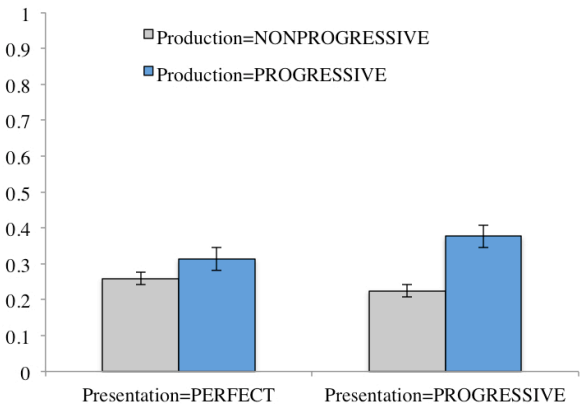


Figure 2. Mean proportion of gestures with iteration according to presentation and production aspect. Error bars show standard error.

To explore this result further, we first looked at the case where presentation aspect was perfect, and compared mean proportion of gestures with iteration in the two production aspects. This test showed no significant difference in the mean number of iterated gestures produced with progressive or non-progressive utterances ($Z = 1.196$, p (two-tailed) = .23). When an event had been originally described in the perfect, people were no more likely to use iterated gestures when they themselves were recounting that event using progressive aspect than when they used perfect aspect.

Second, we looked at the mean proportion of gestures with iteration when presentation aspect was progressive. This test showed a significant difference between the mean rate of iterated gesture for perfect and progressive utterances ($Z = 2.35$, p (two-tailed) = .018). When an event had been originally presented in the progressive, people were more likely to use iterated gestures when they themselves were recounting that event using progressive aspect than when they used perfect aspect. We also compared iteration across production categories. There was no significant difference in mean proportion of gestures with iteration as a function of presentation aspect: $Z = 1.3$; p (two-tailed) = .19 for progressive utterances, and $Z = .66$; p (two-tailed) = .51 for non-progressive utterances.

Finally, we explored the possibility that progressive utterances were accompanied by longer and more iterated gestures because they were simply longer, allowing more time for gesture to be produced. Figure 3 shows mean utterance duration according to presentation and production aspect. When presentation aspect was perfect, we found no significant difference in mean utterance duration as a function of production aspect: $t(31) = 1.22$; p (two-tailed) = .23. That is, progressive utterances were not longer in this case, which argues against an account where progressive utterances simply allow for longer gestures. However, when presentation aspect was progressive, non-progressive utterances were significantly *shorter*: $t(31) = 4.97$; p (two-tailed) < .001. It does not seem to be the case that progressive utterances were particularly long, but rather that non-progressive utterances were particularly short. This is borne out by the near-significant difference in mean utterance duration when non-progressive utterances are compared across presentation categories: $t(31) = 1.92$; p (two-tailed) = .06. Importantly, this result is different from the result seen in gesture. For gesture, duration was roughly the same except when events presented in the progressive were described in the progressive, in which case gestures were longer. For speech, duration was roughly the same except when events presented in the progressive were described with non-progressive aspect, in which case event descriptions were shorter.

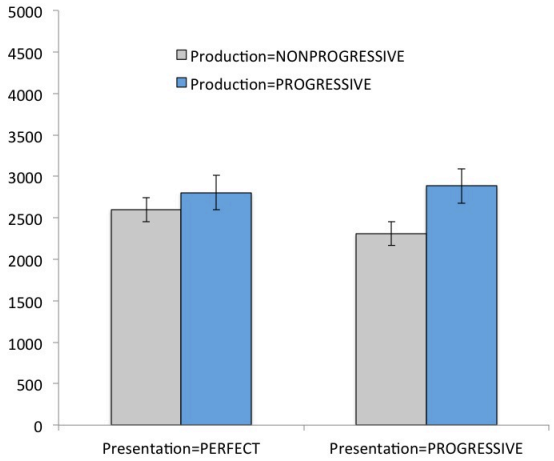


Figure 3. Mean utterance duration (in msec) according to presentation and production aspect. Error bars show standard error.

## Discussion

When people described events using progressive aspect, they produced gestures that were both longer on average and more likely to be iterated, but this occurred only when events had been originally presented with progressive aspect. The interpretation of these results requires some care, as the patterns observed don't appear to arise from either aspect of presentation or aspect of production alone.

To begin with, we have reason to believe that participants did not encode the aspect of presentation very well when events were presented in the perfect. The weakness of this tendency may have been the product of the design, which presented sentences alternating between perfect and progressive. That is, because participants were frequently being asked to comprehend sentences in both aspects, there may have been carry-over effects from one sentence to the next (perhaps in the form of priming) that muddied the results. Had we elected to use a between-participants design, we may have seen stronger effects of presentation alone, for perfect as well as progressive. In addition, the design involved a relatively naturalistic task: telling a story to a friend. The emphasis was thus on being a good communicator, not on perfectly reproducing the stimuli.

The effects of presentation and production aspect on utterance duration were also complex. In general, using the progressive aspect did not appear to make event descriptions longer, but there is no obvious theoretical reason why having an event presented in the progressive and choosing to describe it with non-progressive aspect would make event descriptions *shorter*. Finer-grained analyses of speech might clarify this result.

What might we conclude based on these results? To the extent that this task reflects naturalistic behavior, the results suggest that people don't merely produce longer and more iterated gestures when they're using the progressive; they only do so when they originally learned about the event they're describing using the same aspect. One interpretation of this effect is that congruency between presented and produced aspect is key, perhaps because people encode events with more detail when they're presented with progressive, which allows them the option of recalling them with more vivid detail later on if they choose.

Critically, the results reported here add to those reported in Duncan (2002), showing that the gesture differences between perfect and progressive utterances are not merely the product of different events with different intrinsic properties tending to be described with the perfect or the progressive. We presented participants with the same events in two different aspects. Depending on aspect of presentation and aspect of production, we saw different gestural behaviors.

These results also raise some interesting possibilities for the issue of how simulation during comprehension might interact with simulation during production. If mental simulation occurs during language comprehension, representations formed then should serve as the input for utterances formulated during the process of language production. If simulation also occurs during production, this second process might be congruent with the first, or might conflict with and override it. Our naturalistic narrative task suggests that we are indeed looking at a complex and dynamic process.

## Acknowledgments

Funding was provided by the W.P. Jones Foundation.

## References

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617-645.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society of London: B, 364*, 1281-1289.

Bergen, B., & Chang, N. (2005). Embodied construction grammar in simulation-based language understanding. In J.-O. Östman & M. Fried (Eds.), *Construction Grammar(s): Cognitive Grounding and Theoretical Extensions* (pp. 147-190). Amsterdam: Benjamin.

Bergen, B., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science, 31*, 733-764.

Bergen, B., & Wheeler, K. (2010). Grammatical aspect and mental simulation. *Brain and Language, 112*, 150-158.

Comrie, B. (1976). *Aspect*. Cambridge: Cambridge University Press.

Dowty, D. (1977). Toward a semantic analysis of verb aspect and the English "imperfective" progressive. *Linguistics and Philosophy*, 1, 45-77.

Duncan, S. D. (2002). Gesture, verb aspect, and the nature of iconic imagery in natural discourse. *Gesture, 2*, 183-206.

Glenberg, A., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 96*, 558-565.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gesture as simulated action. *Psychonomic Bulletin & Review, 15*(3), 495-514.

Hostetter, A. B., & Alibali, M. W. (2010). Language, gesture, action! A test of the Gesture as Simulated Action framework. *Journal of Memory & Language, 63*, 245-257.

Kaschak, M. P., & Glenberg, A. M. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory & Language, 43*(3), 508-529.

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., & Zwaan, R. A. (2005). *Perception of motion affects language processing. Cognition, 94*(3), 79-89.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory & Language, 48*(1), 16-32.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Langacker, R. W. (1991). *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Berlin: Mouton De Gruyter.

Madden, C. and Zwaan, R. (2003). How does verb aspect constrain event representations? *Memory & Cognition*, 31, 663-672.

Matlock, T. (2004). Fictive motion as cognitive simulation. *Memory & Cognition, 326*, 1389-1400.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.

McNeill, D. (2005). *Gesture and Thought*. Chicago: University of Chicago Press.

Zwaan, R. A. (1999). Embodied cognition, perceptual symbols, and situation models. *Discourse Processes, 28*(16), 81-88.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 136*, 168-171.