DE GRUYTER MOUTON
DOI 10.1515/cog-2013-0005 — Cognitive Linguistics 2013; 24(1): 135–158

Fey Parrill, Benjamin K. Bergen and Patricia V. Lichtenstein

Grammatical aspect, gesture, and conceptualization: Using co-speech gesture to reveal event representations*

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 was writing a book*) is thought to reflect increased focus on the internal details of an event, as contrasted with the perfect (*I had written a book*). However, experimental evidence that speakers describing events using progressive versus non-progressive aspect are in fact thinking about the same events differently is lacking. We used co-speech gesture as a means to investigate what speakers' event representations are like when they produce progressive versus non-progressive utterances. We found that progressive event descriptions were accompanied by longer-lasting and more complex gestures, but only when participants described events originally presented to them in the progressive. This evidence suggests that people are actually construing events differently when they use different grammatical aspects, but that the aspect originally used to encode the events plays a role as well.

Keywords: gesture, grammatical aspect, language production, mental simulation

Fey Parrill: Department of Cognitive Science, Case Western Reserve University 10900 Euclid Ave, Cleveland, OH 44106. E-mail: fey.parrill@case.edu
Benjamin Bergen: Department of Cognitive Science, University of California, San Diego
Patricia Lichtenstein: Department of Cognitive and Information Sciences, University of California, Merced

1 Introduction

Imagine that a colleague is recounting the thrilling tale of her escape from a mountain lion, and says one of the following two things, *I was running up the path*

* This project was supported by the W.P. Jones Foundation at CWRU. We thank Kensy Cooperrider, Tyler Marghetis, Rafael Núñez, Teenie Matlock and two anonymous reviewers for feedback.

or *I had run up the path*. These descriptions are linguistically different in a crucial way: In the first case, the speaker uses the past progressive (*was running*), while in the second she uses the past perfect (*had run*). Does this linguistic difference reflect any real difference in how she is construing the event? Linguists have long argued that the use of progressive forms indicates a finer-grained construal of the event itself, one in which the speaker is focused on the internal structure of the event. However, while experimental support for such claims in the domain of language comprehension is accumulating, evidence in the domain of language production is harder to come by.

Now imagine that while the same speaker says *I was running up the path*, she produces a gesture showing the curvy, upward trajectory of her escape route, simultaneously wiggling her fingers to show the motion of her legs. Such a gesture provides a great deal of information about the internal structure of the event, as contrasted with a gesture in which an index finger simply traces a straight trajectory. The use of a more complex gesture provides an independent indication that the speaker has adopted a fine-grained construal of the internal structure of the event (Duncan 1996, 2002; McNeill 2003). Thus, if claims about the function of the progressive are correct, it might be reasonable to expect longer, more complex gestures to occur more frequently with progressive utterances than with non-progressive ones. In this study, we probe the relationship between grammatical aspect and the semantic properties of gesture. We ask whether speech-accompanying gestures convergently support the claim that speakers producing progressive utterances are mentally focusing on event-internal structure.

2 Aspect and gesture

Most of the world's languages mark the structure of events, using the linguistic device known as *grammatical aspect*. There are a variety of attested aspectual distinctions, 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 to be conceived of 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 degree of detail with which people think of events (e.g., Comrie 1976; Dowty 1977; Langacker 1991). This intuition jibes well with

native speaker intuition, but can we find empirical support for it in multimodal language?

3 Aspect in language comprehension

Recent work addresses the effects of different grammatical aspects on language comprehension. Several studies (Madden and Zwaan 2003; Bergen and 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 and 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 and Kaschak 2002) with progressive sentences. The action-sentence compatibility effect is a phenomenon wherein participants are faster to respond during a sentence judgment task when the action performed to make that judgment is compatible with the implied meaning of the sentence. For example, moving the arm away from the body to push a button indicating that the sentence *Mark closed the drawer* makes sense is a compatible action and sentence combination, while moving the arm towards the body to push a button indicating the same thing would be an incompatible combination. In Bergen and Wheeler (2010), 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 represent the nucleus of an event (in this case, a motor action) with greater detail than the perfect does.

A second body of experimental research (e.g., Fausey and Matlock 2011; Matlock 2011; Matlock et al. in press) has investigated the role of grammatical aspect in motivating the conceptual shape of events. Matlock (2011) measured the amount of action participants conceptualized in response to a progressive cue (*When John was walking to school*) versus a non-progressive cue (*When John walked to school*). She found that participants were more likely to produce multiple-action main clauses in response to the progressive introductory clause. In two subsequent studies, she found that participants' estimates of the amount of action to have occurred, in terms of both content and temporal extent, were higher in response to progressive sentences than in response to non-progressive sentences. She interpreted these results as indicating that grammatical aspect can facilitate inferences about the amount of action associated with a particular

event. Fausey and Matlock (2011) found that grammatical aspect, in particular the progressive aspect, can amplify the effects of negative information on people's evaluations of propositions and can influence their subsequent reasoning about the predicted outcomes. A study of aspectual framing (Matlock et al. in press) provided further evidence that the progressive aspect can increase the focus on salient information during conceptualization. In this study, participants were prompted to describe a recently viewed event using a progressive or a non-progressive cue. Participants who were prompted with a progressive cue produced more motion verbs and more iconic gestures. In addition, Anderson et al. (2010) used mouse-tracking as a measure of participants' interpretations of progressive and non-progressive forms. The authors found, among other results, that participants' mouse trajectories took more time to produce for progressive forms. Taken together, these studies support the general claim that grammatical aspect modulates conceptualization.

The finding that aspect modulates what part of the mental representation of an event comprehenders focus on has been interpreted in terms of differences in the "mental simulations" constructed by comprehenders. The mental simulation hypothesis proposes that understanding language involves generating perceptual and motor representations of what is being talked about. These mental simulations are believed to be used for different components of language comprehension, such as access to lexical semantics or generation of inferences. Substantial theoretical and empirical work now supports this view (Barsalou 2008, 2009; Bergen 2012; Bergen and Chang 2005; Bergen et al. 2007; Bergen and Wheeler 2010; Glenberg and Kaschak 2002; Hostetter and Alibali 2008, 2010; Kaschak and Glenberg 2000; Kaschak et al. 2005; Matlock 2004; Matlock 2010; Zwaan 1999; Zwaan et al. 2002). For this reason, 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.

Thus, evidence does exist to show that reading or hearing progressive sentences encourages a language user to simulate the internal details of an event. There has been far less work on aspect and the mental operations of people *producing* language (though see Matlock et al. in press, for an interesting exception). Simulation-based theories usually assume that similar principles operate during comprehension and production (e.g., Barsalou 2009; Bergen and Chang 2005), but uncovering the operating principles that underlie language production is difficult. One potentially informative indicator can be found in 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 and Özyürek 2003), but recent work has begun to link research on gesture specifically to simulation-based theories of language (Hostetter and Alibali 2008, 2010; Parrill 2010a). Hostetter and Alibali (2008, 2010) have suggested that gestures might reflect simulations carried out during language production. They propose a model known as the Gestures as Simulated Action framework (GSA) wherein perceptual and motor representations automatically become active during language production and, under certain conditions, are sources of gestures. Gestures, then, reflect the underlying representations associated with a particular utterance. In the present study, gesture may shed light on whether the aspectual choices made by speakers reflect simulation differences during language production.

4 Aspect and gesture in language production

The possibility that gestures accompanying progressive-marked speech reflect greater event-internal focus was first raised in work by Duncan (1996, 2002), and is echoed in a qualitative study by McNeill (2003). Duncan (2002) compared Mandarin Chinese and English descriptions of video stimuli to determine whether gestures varied according to the aspectual form used in speech. She observed that aspectual forms indicating focus on the internal structure of an event (progressive and *durative*, an aspectual category indicating temporal extension) were accompanied by longer gestures when compared to those indicating a construal of the event as completed (*perfective* forms). She also found that progressive-marked utterances in particular were more likely to be accompanied by gestures that were complex, typically including some kind of iterated or repeated gestural motion. Increased complexity can indicate a focus on the internal structure of an event.

The current study uses gesture and language to assess the extent to which produced aspects reflect different construals of the described event. In doing so, it complements existing work on how aspect affects event construal in comprehenders by addressing the relation between aspect and language production in speakers. The study also provides systematic support for the claim that gesture can be used as a window onto event construal: We examine a much larger and more carefully controlled dataset than in previous studies (Duncan 2002; McNeill 2003).

5 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 may take longer to produce and they may encode additional semantic information about 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 (e.g., *was floating*), and half in the past perfect (e.g., *had floated*). We then asked participants to describe the events depicted in the texts to a listener, a task that tends to elicit gesture. This design allowed us to look at two things. First, we were able to 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 were able to 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 are subsequently accessed during language production. In this way, 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 progressive aspect in production might correlate with more complex and longer lasting gestures. This correlation could occur regardless of whether presentation affects production.

5.1 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 a motion event likely to evoke gesture. 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, so that we would have data for each event in both aspects. That is, in the A version, the first, third, and fifth target events appeared in the past progressive, while in the B version those same events 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 story characters *had done*). We used past progressive and past perfect (unlike the present progressive and present perfect found in Bergen and Wheeler (2010), for instance) because the past tense is more typical for story narratives. We did not use simple past because it is aspectually ambiguous, whereas the past perfect is aspectually marked.[1] A sample text is found below. Target event verbs are in bold. The full set of stories can be found in the appendix.

> 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 who did not participate in the main experiment 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, who also did not participate in the main experiment, 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: $t(19) = .86$; $p = .40$.

5.2 Participants and procedure

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 the space bar to advance to the next sentence. After the final sentence, the participant was prompted to move over to the recording area and to

1 One might wonder whether the past perfect is particularly infrequent, and thus particularly unnatural. We were unable to find a comparison of the relative frequency of occurrence of past progressive and past perfect for narrative speech. Biber (2005) has shown that for both conversation and fiction, simple forms dominate, with both progressive and perfect being infrequent. Thus, while past perfect may not be common, we don't have evidence that it is *less* common than past progressive.

describe the story to his or her partner. The partner came to the study with the participant and served as a listener. (Having participants narrate to a friend tends to evoke more naturalistic narrations: Parrill 2010b.) 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).

5.3 Coding

All coders were blind to experimental condition. 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. Non-progressive utterances were typically simple past, but this category also included simple present and past perfect. Because our hypotheses concerned the progressive specifically, we did not choose to make any finer distinctions for non-progressive aspectual forms. One coder carried out this analysis for the entire dataset, and a second coded 42% of the dataset.[2] 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 (using both audio and video) according to the following categories (after McNeill 1992, 2005): *concrete iconic* (the shape and/or motion of the hands maps onto some aspect of the scene being described), *metaphoric iconic* (the shape and/or motion of the hands maps onto an abstract property, such as discourse content), *deictic* (a pointing gesture), *beat* (a rhythmic gesture with no clear 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 *iteration*. Iteration is a repeated action of the hands and one of the ways in which a gesture can exhibit complexity. 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

2 For all reliability coding, the second coder coded a randomly-selected subset of the participants from the complete dataset. Coding 20–25% of the dataset is typical in gesture research where reliability coding is done.

encode the internal structure of the event, by having the fingers wiggle while the hand moves in a straight line. The second gesture would be considered more complex, as it encodes via iterated movement an additional semantic feature (in this case, manner of motion: see Duncan 2002; and see Kita and Özyürek 2003; Parrill 2011, 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. 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 30th 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, the gesture was coded for the presence of iteration, and the duration of that gesture's stroke phase was calculated. Below, we provide examples of gestures illustrating the contrast we are particularly interested in. In Figure 1, the speaker produces a complex (iteration present) gesture accompanying progressive speech, while in Figure 2, the speaker produces a simple gesture (no iteration present) with non-progressive speech.

5.4 Results

Participants produced a total of 960 concrete iconic gestures, with a mean of 30 gestures (SD = 11.94) and 362.75 words (SD = 143.22) per participant. The mean gesture rate (the number of gestures divided by number of words for each participant) was .10 (SD = .09). Our remaining results will be presented in terms of

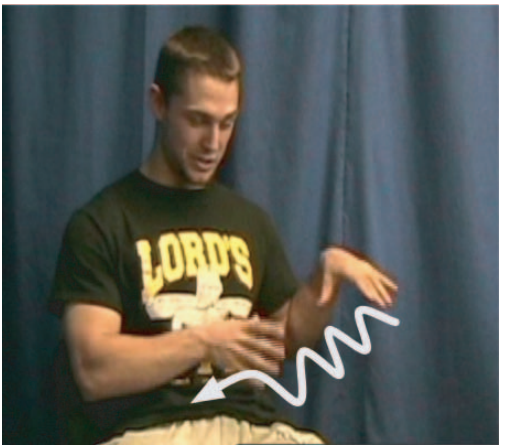

Fig. 1: Speech: *The stick was floating downstream*. Gesture: Both hands move down and to the right with superimposed up and down motions

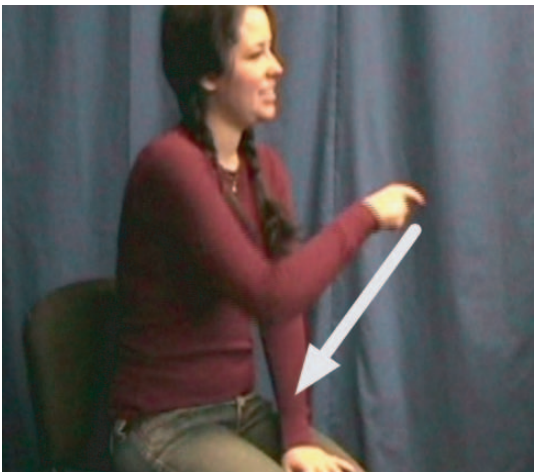

Fig. 2: Speech: *It had floated down*. Gesture: Right index finger traces path downward and to the right

presentation aspect (the aspect that stimulus sentences appeared in) and *production aspect* (the aspect of participants' utterances). Because we controlled the aspect participants received in the text, but did not control the aspect they chose to use in their descriptions, these two factors had the potential to relate in various ways. For example, participants who read a sentence in the progressive might use

the progressive or the past perfect or something else entirely in their descriptions of that sentence. Because of this lack of control over the relation between the aspect a speaker read and the aspect a speaker produced, our first set of analyses focuses on this relationship and not on our main research questions about gesture duration and complexity.

5.5 Presentation aspect and production aspect

First, we asked whether presentation aspect affected production aspect in general, regardless of whether a gesture was produced. We analyzed this relationship by considering the extent to which participants produced utterances where production aspect matched presentation aspect. However, because participants produced different numbers of utterances describing target events from the stimuli, we examined the *proportion* of matches for each participant (where a match is an utterance where presentation and production aspect were the same), rather than just the frequency of matches. Table 1 shows these proportions (SDs in parentheses). Recall that because our hypotheses concerned the progressive specifically, we did not choose to make any finer distinctions for non-progressive aspectual forms, thus the categories shown are progressive and non-progressive.

Participants did not always perfectly reproduce what they read, which is not surprising given that the design had them read sentences that alternated within a passage between perfect and progressive. A Wilcoxon Signed Ranks test showed that participants were not significantly more likely to produce non-progressive when presentation aspect was perfect: $Z = .97$, $p = .32$.[3] However, when presentation aspect was progressive, participants were significantly more likely to produce progressive: $Z = 3.41$, $p < .01$. We next examined the proportion of matches for each participant when a concrete iconic gesture was produced along with an utterance. These means were identical to those in Table 1, so are not shown separately. Again, a Wilcoxon Signed Ranks test showed that participants were not

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

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

3 Two-tailed p values will be reported throughout.

significantly more likely to produce non-progressive when presentation aspect was perfect: $Z = 1.16$, $p = .25$. However, when presentation aspect was progressive, participants were significantly more likely to produce progressive: $Z = 3.32$, $p < .01$. 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 ($Z = .35$, $p = .73$). We also assessed whether participants were more likely to gesture when producing progressive or non-progressive utterances. There was no significant difference in the mean proportion of gestures produced as a function of aspect in the accompanying speech ($Z = 1.62$, $p = .10$).

5.6 Presentation, production aspect and gesture

We then turned to our main research questions. First, a brief comment on statistics is needed: Presentation aspect in this design is an independent variable but production aspect is a *dependent* variable. That is, while we controlled aspect of presentation, aspect of production was, as demonstrated above, affected by presentation aspect. For this reason, a two-factor ANOVA is not appropriate for these analyses (assumptions of independence are violated). When looking at properties of gesture as dependent measures, we report results of Wilcoxon Signed Ranks tests for non-parametric data for frequencies, and paired t-tests for durations.

Our first analysis focuses on the relationship between presentation aspect, production aspect, and gesture stroke duration. Figure 3 shows the mean gesture stroke duration for utterances according to presentation and production aspect.

Overall, gesture stroke duration was longer when the aspect of production was progressive: $t(31) = 2.04$; $p = .05$. This confirmed the principal hypothesis that longer-lasting gestures would accompany the production of progressive utterances. To explore this result further, we first looked at only the cases where presentation aspect had been perfect, and compared mean gesture stroke duration when production aspect was progressive versus non-progressive. This test showed no significant difference between these means ($t[31] = .87$; $p = .38$). That is, when an event was originally presented using the perfect, participants' gestures were not significantly longer when they themselves were recounting that event using progressive aspect, as compared to when they used non-progressive aspect. To make this point more concrete, when a participant read a sentence like *A woman*

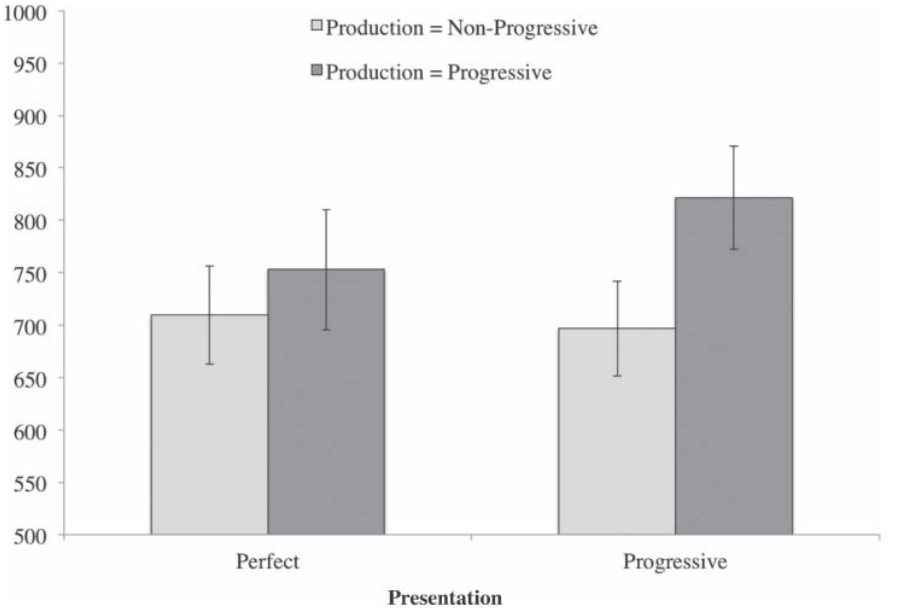

Fig. 3: Mean gesture stroke duration (in msec) according to presentation and production aspect. Error bars show standard error

had hiked down a little hill on the winding path, the participant's gestures were not different in duration on average regardless of whether he or she said *had hiked* or *was hiking* when describing the event.

We then looked at gesture stroke duration when presentation aspect had been progressive, again comparing progressive and non-progressive production aspect. This test showed a marginally significant difference between these means ($t[31] = 2.00$; $p = .05$). When an event was originally presented using progressive aspect, participants made longer-lasting gestures when they themselves were recounting that event using progressive aspect as compared to when they used non-progressive aspect. Again, an example may provide clarity. When a participant read a sentence like *A woman was hiking down a little hill on the winding path*, the participant's gestures were longer on average when he or she said *was hiking* when describing the event, as compared to when he or she said *had hiked*.

We also compared gesture stroke duration *across* presentation categories. That is, the tests above compare progressive utterances to non-progressive utterances within a presentation category, while the following tests compare progressive to progressive across presentation categories, and non-progressive to non-progressive across categories. There was no significant difference in duration as a

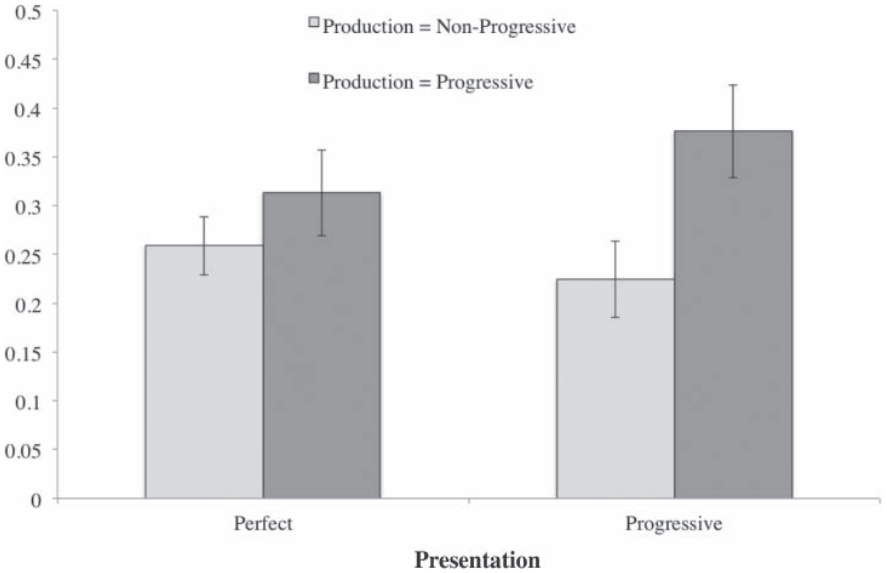

Fig. 4: Mean proportion of gestures with iteration according to presentation and production aspect. Error bars show standard error

function of presentation aspect for progressive utterances ($t[31] = 1.14$; $p = .26$) or for non-progressive utterances ($t[31] = .29$; $p = .77$). In other words, whether a participant read *was hiking* or *had hiked* did not appear to have an effect on gesture stroke duration. These comparisons highlight the fact that the effect we observe is not one of presentation aspect alone.

We then turned to our second dependent measure, gesture iteration. Figure 4 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: $Z = 2.32$, $p = .02$. This confirmed the central hypothesis that progressive utterances would be accompanied by more complex gestures.

To explore this result further, we first looked at the case where presentation aspect had been perfect, and compared mean proportion of gestures with iteration across the two production aspects. This test showed no significant difference in the mean number of iterated gestures produced with progressive or non-progressive utterances ($Z = 1.20$, $p = .23$). When an event had originally been presented in the perfect, people were no more likely to use iterated gestures when they themselves were recounting that event using progressive aspect as compared to when they used non-progressive aspect.

We also looked at the mean proportion of gestures with iteration when presentation aspect had been progressive. This test showed a significant difference between the mean rate of iterated gesture for perfect and progressive utterances ($Z = 2.35$, $p = .02$). When an event had originally been presented in the progressive, people were more likely to use iterated gestures when they themselves were recounting that event using progressive aspect (as compared to when they used non-progressive aspect).

As with duration, we also compared iteration across presentation categories. There was no significant difference in the mean proportion of gestures with iteration as a function of presentation aspect for progressive utterances ($Z = 1.3$; $p = .19$) or for non-progressive utterances ($Z = .66$; $p = .51$). Again, these comparisons highlight the fact that the effect we observe does not arise from presentation alone: It does not appear to be the simple fact of reading *was hiking* (as compared to *had hiked*) that causes participants to include iteration in gesture more frequently.

6 Discussion

When people described events using progressive aspect, they produced gestures that were both longer on average and more likely to be iterated. However, a closer look revealed that this effect only occurred reliably when events had been originally presented with progressive aspect. The interpretation of these results requires some care. What is it about reading a progressive utterance that might have made participants more likely to produce a longer, more complicated gesture when they chose to use the progressive?

In interpreting this effect, we have to deal with the fact that all produced progressive utterances do not behave equivalently with respect to gesture. This means, concretely, that to the extent that longer-lasting and more iterated gestures are indicative of increased mental focus on the internal structure of the event, only some progressive utterances indicate increased mental focus on the internal structure of the event. Namely, these are utterances describing events that speakers learned about originally in the progressive. So we can rephrase our question from above in more specific terms. Why would speakers focus more on event-internal structure when producing progressive utterances, but only when they learned about those utterances in the progressive?

We know from previous work on language processing that when comprehenders read or hear progressive utterances, they encode more event-internal detail of the described events. So participants in the experiment were likely to have encoded more event-internal detail for events presented in the progressive than

for those presented in the perfect. When they subsequently described the event, speakers who read the progressive originally, and therefore had detailed event-internal representations of the events easily available, were more likely to produce progressive utterances, and when they did so, to produce longer-lasting and iterated gestures. However, when speakers had originally read about an event described in the perfect, they had less detailed event representations readily available for production, and as a result, the progressive utterances they produced did not show gestural signatures of increased event-internal representation.

A straightforward but theoretically interesting consequence of this interpretation of the observed effects is that aspect may not always indicate mental construal of event structure. If the reasoning we've presented holds, it follows that only when people have sufficient event-internal detail available does their use of progressive aspect indicate increased focus on the internal structure of events. In a way, this wouldn't be an outlandish claim. Because aspect is obligatorily marked in so many linguistic contexts, it's possible that it sometimes does and sometimes does not reflect underlying speaker conceptualization.

This line of reasoning also makes a testable prediction. We've argued that production aspect might not always indicate what speakers are focusing on, only when they actually have encoded the event in sufficient detail. If this is right, then the criterion of sufficient detail should apply not only to the internal structure of events but also the consequences and endstates of events. So if one were to develop a gestural measure indicating that speakers are focusing more on event endstates (possibly involving the relative forcefulness of a gesture's final movement prior to retraction), then when speakers first encoded the event as readers in the perfect, they should show more of an increase in this measure with produced non-progressives than when they encoded the event in the progressive. To our knowledge, this prediction has not been tested in the literature.

It is also interesting to consider these findings in relation to claims of the Gestures as Simulated Action (GSA) framework (Hostetter and Alibali 2008, 2010). This model predicts that the probability of a gesture being produced depends on several factors, including the strength of activation of a motor simulation. The GSA framework does not have anything to say about whether a gesture is likely to be long or short, or whether it is likely to be complex or simple. However, our work suggests that properties of a simulation predict properties of a gesture, a finding that is generally in line with the GSA claims. Indeed, Hostetter and Alibali (2008) speculate that point of view in gesture may indicate differences in simulation, suggesting that the framework could be extended to include properties such as duration and complexity.

While the patterns we observed in gesture are complex, they are helpful in ruling out one alternative explanation for our principal gesture results, namely

that progressive *utterances* were simply longer in duration, allowing more time for gesture to be produced. That is, if it takes longer to say *was hiking* than *had hiked*, there might be more time to produce gesture, and those gestures might have been longer and more complex. Because gestures accompanying progressive utterances were not *always* longer (we only saw this effect when aspect of presentation was progressive), we do not find this explanation compelling. However, to rule it out more conclusively, we calculated utterance duration for all target utterances. A comparison of the mean duration of progressive and non-progressive utterances indicated that progressive utterances were indeed longer ($t[31] = 3.69$; $p = .01$). However, a closer inspection suggests that the differences in utterance duration do not coincide with the differences in gesture duration. In the case of gesture, progressive-accompanying gestures were in general longer than non-progressive-accompanying gestures due to an increase in the duration of progressive-accompanying gestures across presentation categories. In the case of speech, progressive utterances were in general longer than non-progressive utterances due to a decrease in the duration of non-progressive utterances. In other words, although progressive utterance duration remained largely stable across presentation categories, progressive-accompanying gestures were longer for the same set of utterances. In addition, although non-progressive utterance duration was shorter across presentation categories, non-progressive-accompanying gesture duration remained largely stable across presentation categories. If gesture duration were closely coupled with utterance duration, measures of the two could be expected to follow similar trends, which is not what we found. Figure 5 shows the mean utterance durations across categories.

Indeed, we found a near-significant difference for non-progressive utterances ($t[31] = 1.92$; $p = .06$) when compared across presentation categories: non-progressives were shorter when presentation aspect had been progressive.[4] There is no obvious theoretical reason why this effect would occur. Finer-grained analyses of speech (that is, analyses of different speech units, such as verb phrases rather than entire utterances) might clarify this result.

The properties of our stimuli suggest several other possibilities for future research. First, in claiming that the progressive indicates focus on the internal structure of events, we gloss over a great deal of complexity about the different

4 In addition, progressive utterances were not longer than non-progressives when presentation was perfect ($t[31] = 1.22$; $p = .23$), which argues against an account where progressive utterances simply allow for longer gestures. Progressive utterances did not differ across presentation categories ($t[31] = .42$; $p = .67$), and while there was a significant difference in mean duration within the progressive presentation category ($t[31] = 4.97$; $p < .001$), again, this appears to arise from non-progressives being unusually short in that condition.

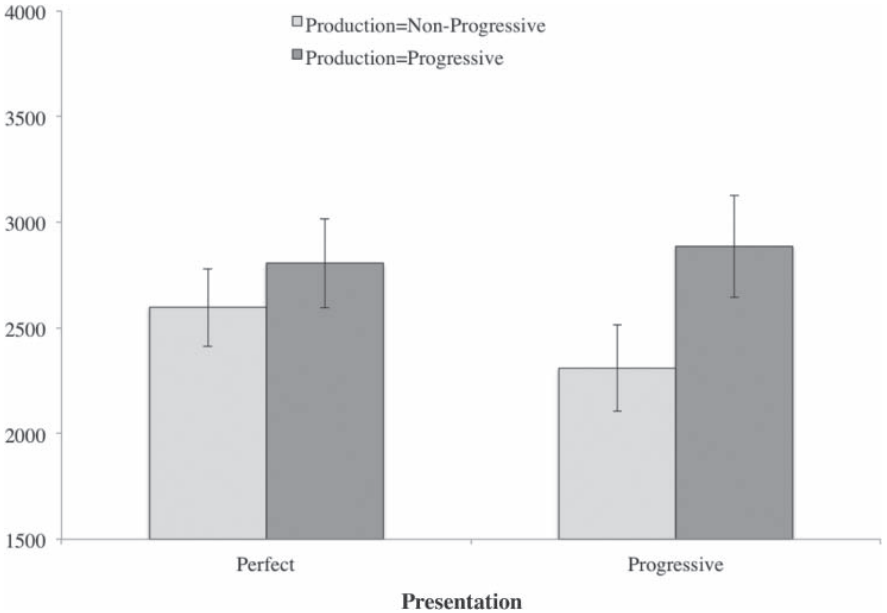

Fig. 5: Mean utterance duration according to presentation and production aspect. Error bars show standard error

kinds of events that can appear in the progressive. Vendler (1967) has fruitfully classified events as *states* (situations that don't "go on in time" (1967: 99), but are rather true or false at a particular moment, such as *know*), *achievements* (situations that occur at a single moment, as in *reach the top*), *activities* (situations with internally homogeneous structure, such as *walking*), and *accomplishments* (situations with an inherent boundary, such as *drawing a circle*). While a full discussion of the subject is beyond the scope of this paper, it is noteworthy that our target events contain activity and accomplishment verbs. As pointed out by a reviewer, it is entirely possible that achievements (more likely to be infelicitous in the progressive, but not impossible, as in *he was reaching the top*) might pattern differently, or that there might be interesting gestural differences between our activity and accomplishment events.[5]

5 In addition, while we carefully held constant some features of our stimuli (animacy, humanness) there was more variability in the *number* of the syntactic subjects of our target events: Some were singular, some plural. As pointed out by a reviewer, number has consequences for the interpretation of aspectual marking, so it is possible that further analyses might have revealed patterns of gestural behavior as a function of plurality. We chose not to

Second, the gestures we analyze all convey semantic information about path and manner of motion. While motion is a natural place to explore the relationship between aspectual form, gestural form, and underlying construal, it is conceivable that our findings would not generalize beyond the domain of motion events. An interesting extension of this study would be to consider cases in which the semantic information present in gesture is metaphorical, rather than iconic. That is, if we were to see similar effects for utterances that do not involve actual motion, such as psychological state predicates (*thinking*, *feeling*, etc.), we might have a stronger argument for the claim that gesture provides information about event construal. We do know that speakers produce gestures involving iteration for such metaphorical expressions (Cienki 1998; Sweetser 1998), but to date no experimental work has compared gesture production that accompanies metaphorical language with different aspects.

7 Conclusion

Speakers produce longer-lasting and more complex gestures when producing progressive utterances than when producing non-progressive ones. This suggests that in language production, as has been shown previously for language comprehension, aspect is an indicator of the parts of an event that people focus on. These findings add to those reported in Duncan (2002), by showing that differences in gestures accompanying progressive and non-progressive utterances can be observed with a carefully controlled set of events, and over a larger group of narrators.

However, our results also showed that speakers only produced longer and more iterated gestures with progressive aspect when they had originally learned about the event with progressive aspect. There are two clear implications of this latter finding.

The first implication is that the longer-lasting and more iterated gestures we observed with progressive utterances weren't just due to progressive utterances lasting longer as a whole. Progressive utterances that described events originally learned about through perfect aspect showed no significant increase in gesture length or iteration. This eliminates a possible confounding factor that could otherwise explain the central results.

---

analyze each target event separately (item analyses), because different participants described different numbers of target events, leading to differences in the number of data points for each item. As a result, we are unable to provide an answer to this particular question, but it would be interesting to follow up on this possibility.

The second implication relates to the broader interpretation of the results. Because gesture duration and iterativity increased only for progressive utterances that were originally presented in progressive, the fact that a speaker uses one aspect or another doesn't necessarily tell us anything definitive about his or her conceptualization of the described event. Our results showed that production of the progressive goes along with increased focus on event-internal structure, but only when we know that the speaker learned about the event in such a way that he or she was encouraged to encode the event-internal structure (i.e., because the original report of the event was given using the progressive).

Finally, while the specific findings of this study contribute to the literature on aspect, we believe this work also makes a more general point about the nature of language. Researchers who study speech and gesture in conjunction have long argued that language involves both an imagistic-motoric component, and a symbolic component (see McNeill 1992, for early versions of this proposal). This general claim is highly compatible with the tenets of cognitive linguistics, and with psycholinguistic work falling under the umbrella of mental simulation (wherein using language involves the generation of perceptual and motor representations). However, we are still a long way from being able to specify how processes of simulation or imagery-generation during comprehension relate to processes of simulation during speech and gesture production, and how such process should be integrated into a model that provides testable predictions. Studies such as ours shed light on the interaction between specific grammatical devices and underlying conceptualization, and can thus provide more detailed and systematic support for a multimodal model of language.

References

Anderson, Sarah, Teenie Matlock & Michael Spivey. 2010. The role of grammatical aspect in the dynamics of spatial descriptions. In C. Hölscher, T. Shipley, M. Olivetti Belardinelli, J. Bateman, & N. Newcombe (eds.), *Spatial Cognition*, 139–151. Berlin/Heidelberg: Springer.

Barsalou, Lawrence W. 2008. Grounded cognition. *Annual Review of Psychology*, 59, 617–645.

Barsalou, Lawrence W. 2009. Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society of London: B*, 364, 1281–1289.

Bergen, Benjamin. 2012. *Louder than words: The new science of how the mind makes meaning.* New York: Basic Books.

Bergen, Benjamin, & Nancy Chang. 2005. Embodied construction grammar in simulation-based language understanding. In J.-O. Östman & M. Fried (eds.), C*onstruction Grammar(s): Cognitive Grounding and Theoretical Extensions*, 147–190. Amsterdam: Benjamin.

Bergen, Benjamin, Shane Lindsay Teenie Matlock & Srini Narayanan. 2007. Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science*, 31, 733–764.

Bergen, Benjamin, & Kathryn Wheeler. 2010. Grammatical aspect and mental simulation. *Brain and Language*, 112, 150–158.

Biber, Douglas. 2005. Corpus linguistics and the study of English grammar. *Indonesian Journal of English Language Teaching*, 1(1), 1–22.

Cienki, Alan. 1998. Metaphoric gestures and some of their relations to verbal metaphoric expressions. In J.-P. Koenig (ed.), *Discourse and cognition: Bridging the gap* 189–204. Stanford: CSLI Publications.

Comrie, Bernard. 1976. *Aspect*. Cambridge: Cambridge University Press.

Dowty, David. 1977. Toward a semantic analysis of verb aspect and the English "imperfective" progressive. *Linguistics and Philosophy*, 1, 45–77.

Duncan, Susan. D. 1996. *Grammatical form and 'thinking-for-speaking' in Mandarin Chinese and English: an analysis based on speech-accompanying gestures* (unpublished doctoral dissertation). University of Chicago, Chicago.

Duncan, Susan D. 2002. Gesture, verb aspect, and the nature of iconic imagery in natural discourse. *Gesture*, 2, 183–206.

Fausey, Caitlin M. & Teenie Matlock. 2011. Can grammar win elections? *Political Psychology*, 563–574.

Glenberg, Arthur & Michael P. Kaschak. 2002. Grounding language in action. *Psychonomic Bulletin and Review*, 96, 558–565.

Hostetter, Autumn B. & Martha W. Alibali. 2008. Visible embodiment: Gesture as simulated action. *Psychonomic Bulletin and Review*, 15(3), 495–514.

Hostetter, Autumn B. & Martha W. Alibali. 2010. Language, gesture, action! A test of the Gesture as Simulated Action framework. *Journal of Memory and Language*, 63, 245–257.

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

Kaschak, Michael P., Carol J. Madden, David J. Therriault, Richard H. Yaxley, Mark Aveyard, Adrienne Blanchard, & Rolf A. Zwaan. 2005. Perception of motion affects language processing. *Cognition*, 94(3), 79–89.

Kita, Sotaros & Asli Özyürek. 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 and Language*, 48(1), 16–32.

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

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

Madden, Carol & Rolf A. Zwaan. 2003. How does verb aspect constrain event representations? *Memory and Cognition*, 31, 663–672.

Matlock, Teenie. 2004. Fictive motion as cognitive simulation. *Memory and Cognition*, 32(8), 1389–1400.

Matlock, Teenie. 2010. Abstract motion is no longer abstract. *Language and Cognition*, 2, 243–260.

Matlock, Teenie. 2011. The conceptual motivation of aspect. In K. Panther & G. Radden (eds.), *Motivation in Grammar and the Lexicon*, 133–147. Amsterdam: Philadelphia: John Benjamins.
Matlock, Teenie, David Sparks, Justin L. Matthews, Jeremy Hunter, & Stephanie Huette. 2012. Smashing new results on aspectual framing. *Studies in Language* 36(3), 700–721.
McNeill, David. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.
McNeill, David. 2003. Aspects of aspect. *Gesture*, 3(1), 1–17.
McNeill, David. 2005. *Gesture and Thought*. Chicago: University of Chicago Press.
Parrill, Fey. 2010a. Viewpoint in speech-gesture integration: Linguistic structure, discourse structure, and event structure. *Language and Cognitive Processes*, 25(5), 650–668.
Parrill, Fey. 2010b. The hands are part of the package: Gesture, common ground, and information packaging. In S. Rice & J. Newman (eds.), *Empirical and Experimental Methods in Cognitive/Functional Research*, 285–302. Stanford: CSLI Publications.
Parrill, Fey. 2011. The relation between the encoding of motion event information and viewpoint in English-accompanying gestures. *Gesture*, 11(1), 61–80.
Sweetser, Eve. 1998. Regular metaphoricity in gesture: Bodily-based models of speech interaction. *Actes du 16 Congres International des Linguistes 5*.
Vendler, Zeno. 1967. Verbs and times. *Linguistics in Philosophy*, 97–121.
Zwaan, Rolf A. 1999. Embodied cognition, perceptual symbols, and situation models. *Discourse Processes*, 28(16), 81–88.
Zwaan, Rolf A., Robert A. Stanfield & Richard H. Yaxley. 2002. Language comprehenders mentally represent the shapes of objects. *Psychological Science*, 136, 168–171.

Appendix

Stimulus stories

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

2. It was a windy fall day on the quad. A student was running down the curved walkway through the campus. Then the student had rushed off the walkway toward a large building. A squirrel was scampering across the path toward a tree. A little later, the squirrel had scurried up the tree. With a gust of wind, some loose papers were swirling into a pile by the stairs of the big building. The student had now run up the stairs toward the building's entrance.

3. It was midday at the county fair. A band was marching around the main promenade. At one point, a baton had spun up from the drum major's hands and down again. The drum major was twirling around as she caught it. Meanwhile, on the barnyard race track, the pigs had run around a bend. Then the lead pig was leaping over the low hurdle near the finish line. Behind a crowd of spectators, a toddler had clambered up onto his father's shoulders to get a better look at the race.

4. It was a bright winter afternoon. A couple was climbing up the wooden steps along the side of the hill. At the top of the run some kids had jumped into a toboggan. Then the toboggan was zooming off down the ice hill. Off to the side of the run a small child in a snow suit had toddled into a cluster of trees. A hungry-looking chipmunk was scurrying up one of the trees. Then the chipmunk had bounced from branch to branch deeper into the forest.

5. It was a sunny summer's day in the park. A boy was rolling down a hill onto the meadow. Along the side of the meadow a bunny had hopped into a nearby patch of clover. On the meadow a girl playing softball was running to second base. The ball had bounced off to the side toward the clover patch. Then a kid was darting after it. Meanwhile the bunny had leapt up the hill.

6. It was winter and the small lake was frozen. On one end a boy was sliding onto the ice toward the other side. On the other end a figure skater had twirled up into a spin. An ungainly puppy was sprawling around in a circle in the middle of the ice. A hockey player had zigzagged past the puppy. Soon his puck was bouncing away from him toward a snow bank. Then the puppy, now steady on all fours, had bounded after it.

7. It was warm day in the early fall. A man with a fishing pole was walking down an embankment toward a river. His hound had trotted down ahead of him. By the river a woman was climbing into a canoe. Soon the canoe had glided down river. Meanwhile the man was wading into the water. The hound had trotted back up the embankment.

8. There were only a couple of people at the pool on an overcast day. A water bug was hopping onto the water's surface. Then a big kid had cannonballed into the pool. With the force of his landing the water was swaying up and down along the edges of the pool. Now the bug had flitted away. On the other end of the pool a girl was diving off the high dive in an arching descent. Then she had sliced through the water to the bottom of the pool.

9. Early fall was a good time for bike-tripping in the mountains. A pair of cyclists was zigzagging up the steep mountain road. From a cliff top above them a

red-tailed hawk had dived down toward another hawk. Then the two hawks were flying in wide circles over the river valley. Meanwhile the cyclists had spun down from the peak of the mountain road. Some rocks were rolling down onto the road. The cyclists had swerved around them.

10. It was dusk at the river's edge. Fireflies were flitting into the foliage by the water. A girl had climbed up a big tree on the river's edge. Meanwhile her friend was swinging up over the river on a tire swing hanging from the tree. Then the girl on the swing had vaulted off it into the river. Some fish were finning away in all directions from the point of impact. The tire swing had twisted around on the rope.