

Underspecification and Aspectual Coercion

Martin J. Pickering
University of Edinburgh

Brian McElree
New York University

Steven Frisson
University of Birmingham

Lillian Chen
University of Michigan

Matthew J. Traxler
University of California, Davis

In principle, comprehenders might always make immediate commitments to the interpretation of expressions (full commitment) or wait until such decisions are necessary (minimal commitment; Frazier & Rayner, 1990). One interesting case involves decisions about telicity: whether expressions refer to events that are determinate versus indeterminate with respect to an endpoint. Thus, *the insect hopped* is apparently determinate, but continuing with a clause beginning with *until*, in which case *hopped* must be interpreted as an ongoing activity, is possible. Studies using secondary lexical decision and “stop-making-sense” tasks found that comprehenders experienced difficulty with these continuations, compatible with full commitment (Piñango, Zurif, & Jackendoff, 1999; Todorova, Straub, Badecker, & Frank, 2000a, 2000b). However, we report 2 self-paced reading and 2 eye-tracking experiments that indicate readers do not experience any difficulty with these types of mismatches in telicity. We argue that during normal reading, comprehenders do not immediately need to commit fully to the telicity of events and that full commitment may only oc-

cur when processing demands induce immediate decisions. We contrast these results with evidence for full commitment in complement coercions, for example, *began the book* (McElree, Traxler, Pickering, Seely, & Jackendoff, 2001) and other forms of semantic interpretation.

Early psycholinguistic research often assumed that people waited until the end of a clause or sentence before beginning the process of interpretation (e.g., Fodor, Bever, & Garrett, 1974). However, since the 1970s, much research has suggested that people interpret sentences incrementally (e.g., Marslen-Wilson, 1973), and many researchers now believe that all information is interpreted as soon as it becomes available. For example, Just and Carpenter (1980) proposed the immediacy hypothesis, whereby readers fixate a word until they have determined its interpretation and how this interpretation should be integrated with prior context. However, interpretation is multifaceted in that comprehenders may need to determine aspects of word meaning, as well as phrasal, sentential, and discourse properties (quantifier scope, anaphoric relations, etc.). Some components of interpretation possibly may be resolved immediately and others not, or the immediacy of interpretation depends on general factors such as the degree of ambiguity in the sentence so far. Additionally, task demands in experimental settings and the method used to gather data may affect immediacy (e.g., Magliano, Graesser, Eymard, Haberlandt, & Gholson, 1993).

Frazier and Rayner (1990) contrast a *full commitment* account, whereby comprehenders immediately commit to one possible interpretation of an expression with a *minimal commitment* account, and comprehenders delay commitments until they are necessary. Generally we can contrast full commitment with *incomplete commitment*, whereby comprehenders delay at least some aspects of interpretation on occasion. Any account that admits incomplete commitment is incompatible with the immediacy hypothesis because it assumes that incrementality has limits. When comprehenders fully commit to some aspect of meaning but incompletely commit to others, they must construct an underspecified representation of meaning.

Determining limits on incrementality would be informative for theories of comprehension in part because it would help us to separate components of semantic processing: A component of processing that requires full commitment must be disconnected from those that allow incomplete commitment.

In this article, we focus on *aspectual coercion* and what it may tell us about whether comprehenders make full commitments about aspect during comprehension. The grammatical category of *aspect* refers to the temporal structure of an event. Here, the relevant aspectual property concerns whether the event a verb denoted contains a logical endpoint. For example, the event denoted by *hopped* finishes at the point when the hop is complete. Because a hopping event is short, performing a single hop indefinitely is not possible. In contrast, the event *glided* can

continue indefinitely. The hopping event therefore has a logical endpoint and is *bounded* (sometimes called *telic* or *perfective*); whereas the gliding event does not have a logical endpoint and is therefore *unbounded* (sometimes called *atelic* or *imperfective*). Additionally, each part of a gliding event is still a gliding event, but each part of a hopping event is not a hopping event (but merely part of it).

This semantic analysis indicates that a clear distinction can be made between bounded and unbounded events. However, some expressions are ambiguous about the type of event to which they refer. Do comprehenders fully commit to either a bounded or an unbounded interpretation in such cases, or can they incompletely commit until the intended interpretation is clear? Can they, for example, assume that the preceding noun phrase serves as the agent of the act denoted by the verb, and perhaps that the event took place in the past, without making a decision about whether the event is bounded or unbounded? Note that this ambiguous situation will not arise in languages that use explicit morphological markers to express telicity.

ASPECTUAL COERCION

Typically, an unbounded event such as an insect gliding can continue until something happens to interrupt the event. In contrast, a bounded event such as an insect hopping cannot continue indefinitely. Thus, we might expect that (1a) would be possible but (1b) would not:

- (1a) The insect glided effortlessly until it reached the garden.
- (1b) The insect hopped effortlessly until it reached the garden.

However, Piñango et al. (1999) note that both (1b) and (1a) are possible. Whereas (1a) is straightforwardly compatible with this analysis, (1b) does not seem to be because an act of hopping constitutes a bounded event. However, the sentence makes sense because interpreting *the insect hopped* as referring to a series of hopping events is possible. Although a single hopping event is bounded, a series of hopping events is unbounded and is therefore compatible with the meaning of *until*. Piñango et al. argued that the meaning of *hopped* is coerced from referring to a bounded event to an unbounded series of events so that the sentence becomes felicitous.

To investigate this, they had participants listen to sentences such as (1a) and (1b) and simultaneously perform a lexical-decision task on an unrelated word as an index of processing difficulty. At *until*, lexical decision times were longer in (1b) than (1a). They argued that people initially assigned a bounded (single-event) interpretation to *the insect hopped effortlessly*, and were forced to revise their interpretation at *until*. Piñango et al. (1999) argued that the difficulty making the lexical

decision in (1b) was due to the act of coercing *the insect hopped effortlessly* from a bounded to an unbounded interpretation. This explanation would be compatible with the evidence that other expressions, such as *began the book*, which appear to require a different type of coercion, cause processing difficulty (e.g., McElree et al., 2001; Traxler, Pickering, & McElree, 2002); see the General Discussion section. However, at least two other explanations for this finding exist: that semantic reanalysis is difficult and that constructing any iterative interpretation is difficult. For our purposes, if the difficulty is due to either coercion or semantic reanalysis, the results suggest that people do commit to a particular aspectual interpretation rather than underspecifying the representations with respect to aspect.

Todorova et al. (2000a) reported an experiment that indicates that iterative interpretations are not difficult to construct and that comprehenders delay in committing to an aspectual interpretation of a verb until they have processed some of the verb's arguments. They used a stop-making-sense procedure in which participants read sentences such as (2a–d) phrase by phrase and decided whether they made sense in relation to the prior sentential context (e.g., Boland, Tanenhaus, Garnsey, & Carlson, 1995):

- (2a) Even though / Howard sent / a large check / to his daughter / for many years, / she refused to accept his money.
- (2b) Even though / Howard sent / large checks / to his daughter / for many years, / she refused to accept his money.
- (2c) Even though / Howard sent / a large check / to his daughter / last year, / she refused to accept his money.
- (2d) Even though / Howard sent / large checks / to his daughter / last year, / she refused to accept his money.

Todorova et al. (2000a) pointed out that the aspect of an expression is not merely fixed by the verb but rather by the verb in combination with its arguments. Although *sent* appears to refer to a bounded event, the expression *sent large checks* can refer to an unbounded (iterative) event in which checks are sent repeatedly. If readers fully commit to a bounded interpretation of *sent* (i.e., at the verb), we might expect difficulty with *large checks* because they would have to reanalyze. However, they found no indication that readers disliked *sent large checks* (either in terms of reading times or make-sense judgments). This suggests that iterative interpretations per se are not difficult.

Readers might therefore assume the iterative interpretation of the verb phrases in (2b) and (2d), but the bounded interpretation of the verb phrases in (2a) or (2c). The adverbial *for many years* in (2a–d) is only compatible with the iterative interpretation. So if they adopt the bounded interpretation of (2a), they ought to experience difficulty with reanalysis or coercion at this point. Accordingly, Todorova et al. (2000a) found that readers specifically disliked the adverbial in (2a) versus (2b–d), both in terms of reading times and make-sense judgments. They also found

similar effects when readers were subsequently asked to make sense judgments for the entire sentence.

Todorova, Straub, Badecker, and Frank (2000b) reported a second experiment in which (3a–b) replaced (2c–d):

- (3a) Even though Howard sent a large check to his daughter every year, she refused to accept his money.
- (3b) Even though Howard sent large checks to his daughter every year, she refused to accept his money.

Their concern was that some items such as (2a) may have been ambiguous between the intended iterative interpretation and an interpretation in which a single sending event took many years (although this is very unlikely for this example). If so, the difficulty with (2a) may have been due to resolving this ambiguity. This experiment found somewhat different results. Some reading-time difficulty with *large checks* versus *a large check* was found, but the conditions differed in the frequency of the noun in this region (frequency: singular, 38.83 per million according to Francis & Kučera, 1982; plural: 19.36), $t(35) = 3.96$, $p < .001$. Participants spent longer reading the adverbial in (2a) versus (2b) and (3a) versus (3b), but they were much more likely to reject the adverbial in (2a) than any of the other conditions, and no differences were found in the make-sense judgments for the entire sentence. The reading times, although not the rejection rates, therefore suggest that aspectual coercion is costly. However, data from the stop-making-sense task are difficult to interpret unambiguously. Assuming that making a “No” response and switching buttons to make that response take time, the increased reading times can be related to this process rather than regular reading processes. To complicate matters, Proctor, Dickey, and Rips (2004) reported evidence for difficulty when a mismatch occurred between the telicity information in the sentence and a sentence-final adverbial phrase (*for eight minutes* vs. *in eight minutes*). However, this effect did not appear until late in sentence processing and only with comprehension questions that specifically focused on telic interpretations.

Thus, both Piñango et al. (1999) and Todorova et al. (2000a, 2000b) provided some evidence against minimal commitment with respect to aspectual coercion. However, several concerns arise with these conclusions. First, both studies used methods that may not reflect normal language comprehension. Piñango et al. employed a secondary task (lexical decision) whose effects on language comprehension are uncertain. Todorova et al.’s task may persuade people to construct particularly detailed representations because they specifically focus on whether what they are reading makes sense. People may be more likely to underspecify under more standard reading conditions, in particular when the underspecification relates to a comparatively subtle component of interpretation, such as aspectual processing.

Indeed, Todorova et al.'s finding of difficulty at the adverbial but not at the object noun (which Proctor et al., 2004, replicated) suggested that people underspecified while processing the object noun. Perhaps a hard-and-fast rule states that underspecification can only persist during processing of arguments, but more likely the results suggest that underspecification can persist for a length of time and that commitment to a specific aspectual interpretation may depend on multiple factors, including the nature of the task itself. Hence people may automatically determine the aspect of sentences as they encounter them. Equally possible, however, is that they need not resolve aspect unless they make a decision to do so, for example, as a consequence of the task at hand. When people are performing normal reading, they might avoid making such a commitment before it is necessary. Such an account would mean that aspectual underspecification is (to some extent at least) under strategic control with people being more likely to underspecify when they have more processing resources available. This might be a consequence of the complexity of the reading material, the nature of the task, or their level of reading ability.

Indeed, we have good evidence that people do not fully commit with respect to various aspects of semantics and that they might even minimally commit. Frazier and Rayner (1990) found that readers made an immediate commitment to the interpretation of a homonymous noun (e.g., *pupil*), because reading times were longer when context supported the subordinate meaning rather than the dominant meaning. However, they did not find similar effects with polysemous nouns (e.g., *newspaper* meaning object vs. institution), which suggests that readers did not have to select between interpretations. Pickering and Frisson (2001) found that people did not commit immediately to an interpretation of a polysemous verb (e.g., *disarmed*). Even when disambiguating context followed the noun, little sign of difficulty occurred when it supported the subordinate interpretation. Similarly, Frisson and Pickering (1999) found that readers experienced no difficulty with the more or less frequent sense of a word with a literal meaning (e.g., *convent* as building) or a metonymic meaning (*convent* as institution). Frazier, Pacht, and Rayner (1999) found that people experienced difficulty with distributive interpretations such as, "Lyn and Patrick saved \$1,000 each ...," but only when the disambiguating information was delayed. When it occurred immediately after the verb ("Lyn and Patrick each saved \$1,000 ..."), no difficulty ensued. Note that these studies all had participants read series of isolated sentences in which the relevant semantic distinction occurred repeatedly during the experiment. We find, therefore, evidence against full commitment even when the distinctions may be more important than they are in naturalistic discourse. Finally, many observations indicate semantic processing may be incomplete, for example, when interpretations violate presuppositions (e.g., Barton & Sanford, 1993; Ericson & Mattson, 1981; Wason, 1959). Therefore, it seems quite plausible that people need not fully commit with respect to aspect either.

To address these questions, we conducted four experiments, two based on Piñango et al. (1999) and two based on Todorova et al. (2000b). In each case, one experiment used self-paced reading and one used eye tracking. These experiments therefore provide a strong test of whether disambiguating sentences toward an iterative reading causes processing difficulty. If so, they would be compatible with explanations in terms of semantic reanalysis or the difficulty of aspectual coercion (or conceivably difficulty with iteration per se). Additionally, the experiments based on Piñango et al. included conditions in which the adverbial phrase was fronted, so that it occurred before the potentially ambiguous verb (see next). All our experiments also included experiments investigating complement coercion, which causes considerable processing difficulty in many studies (McElree et al., 2001; McElree, Frisson, & Pickering, 2006; McElree, Pykkänen, Pickering, & Traxler, 2006; Pickering, Traxler, & McElree, 2005; Traxler et al., 2002; Traxler, McElree, Williams, & Pickering, 2005). According to these studies, complement coercions such as “The man began the book” cause difficulty because the verb (here, *began*) requires a complement that refers to an event but encounters a complement that refers to an object (here, *the book*). Readers therefore have to introduce new semantic material so that they can interpret the complement as referring to an event involving the object (e.g., reading a book). If the experiments reveal processing difficulty for complement coercion but not for aspectual coercion, they would provide strong evidence that aspect can be underspecified during normal reading.

EXPERIMENT 1: SELF-PACED READING OF PIÑANGO ET AL. (1999)

Experiment 1 asked whether people experienced difficulty during normal reading when they encountered a conflict between a description that appeared to refer to a bounded event such as, “The insect hopped effortlessly,” and a prepositional phrase that was incompatible with this interpretation such as, “... until it reached the garden.” Piñango et al. (1999) found evidence for such difficulty, but their results may have been a consequence of their use of the secondary lexical-decision task. We therefore used two conditions that corresponded to Piñango et al.’s conditions (4a–b), one using a bounded verb such as *hopped*, and the other using an unbounded verb such as *glided*. Additionally, whether any such difficulty would result from having to reanalyze from a bounded to an unbounded (iterative) event or whether any difficulty would simply be due to a clash between the preferred interpretation of the description and the requirements of the prepositional phrase was unclear. To contrast these two possibilities, we included two further conditions in which the prepositional phrase occurred before the description of the event (4c–d). All conditions included a second sentence to increase readability.

- (4a) The insect glided effortlessly until it reached the far end of the garden. It was in a hurry to return to its nest. (unfronted–unbounded)
- (4b) The insect hopped effortlessly until it reached the far end of the garden. It was in a hurry to return to its nest. (unfronted–bounded)
- (4c) Until it reached the far end of the garden, the insect glided effortlessly under the moonlight. It was in a hurry to return to its nest. (fronted–unbounded)
- (4d) Until it reached the far end of the garden, the insect hopped effortlessly under the moonlight. It was in a hurry to return to its nest. (fronted–bounded)

According to Piñango et al.'s (1999) account, readers should experience difficulty in (4b) versus (4a) on or soon after encountering the preposition *until*. If the difficulty is due to the clash between the preferred interpretation of the description and the requirements of the prepositional phrase, readers should experience difficulty in (4d) versus (4c) on or soon after encountering the verb (*hopped* in 4d, *glided* in 4c). If the difficulty is largely or entirely due to the difficulty of semantic reanalysis, then readers should experience little or no difficulty with (4d) versus (4c). However, if readers do not experience difficulty with either (4b) versus (4a) or (4d) versus (4c), then that would suggest that aspectual coercion is not difficult and that people may initially underspecify whether they interpret an event as bounded or unbounded. Note that participants answered comprehension questions after each sentence, which may induce them to process sentences comparatively deeply. In part because naturalistic discourse does not typically involve such questions, our questions were relatively straightforward and did not particularly focus on the aspectual coercion itself.

Method

Participants. Fifty-six native English speakers at New York University participated for credit in an undergraduate psychology course or for monetary compensation. All participants had normal or corrected-to-normal vision.

Stimuli. The stimuli (see Appendix A for sample items) consisted of 24 sets of items such as (4a–d), based on Piñango et al. (1999). We modified the Piñango et al. stimuli such that an additional sentence was added to the end of the original sentences. The unfronted sentences remained in their original form, but the fronted sentences were modified so that the prepositional phrase describing a continuous time period, such as *until it reached the far end of the garden*, preceded the main clause. One version of each item was assigned to one of four lists so that equal numbers of each version appeared on each list and no participant saw more than one version of each item. Each participant was assigned to one of the four lists.

The critical verbs (e.g., *hopped* vs. *glided*) of course differed between conditions, as in Piñango et al.'s (1999) experiment. They note that bounded verbs (e.g.,

hopped) were more frequent than unbounded verbs (e.g., *glided*), and therefore their finding of difficulty with sentences containing bounded verbs could not be due to frequency differences. Because our goal was to provide a comparison with Piñango et al.'s study, this frequency difference also occurred in our experiment.

Procedure. The experimenter instructed participants to read at a normal pace in a manner that would enable them to answer comprehension questions correctly. Sentences were presented using a self-paced moving window procedure built with E-Prime software, an authoring program for creating computerized psychology experiments. Each trial began with a series of dashes on the computer screen in place of the letters in the words. The first press of the “8” on the keypad replaced the first set of dashes with the first region in the sentence. With subsequent presses, the next set of dashes was replaced by the next region, and the preceding region was replaced by dashes. Although most regions consisted of just one word, some regions for the fronted conditions contained two or three words because of specific program requirements. For example, (4a) was presented word by word, but (4c) was presented as follows (slashes indicate regions):

Until / it reached / the / far end / of the garden, / the / insect / glided / effort-
lessly / under / the / moonlight. / It / was / in / a / hurry / to / return / to / its /
nest./

A relatively straightforward yes-or-no question followed each sentence (see Appendix A for sample questions), and participants received feedback on their answers after every trial along with a cumulative accuracy percentage. The computer recorded the time to the nearest millisecond from when a word was first displayed until the next press of the keypad. All participants who contributed to the reported analyses scored 85% accuracy or greater on the comprehension questions.

Results and Discussion

Prior to data analysis, trials in which participants answered comprehension questions inaccurately were excluded from analyses. In total, 3.3% of the data were excluded. Responses were less than 200 ms were recoded as 200 ms, and responses that were greater than 2,000 ms were recoded as 2,000 ms.

We conducted two sets of analyses on the data (see Table 1 for the means). First, we contrasted the time spent reading the verb in the four conditions and the time spent reading the preposition in the four conditions. We then considered each pair of conditions (verb or preposition) separately and looked at the time spent reading the disambiguating word and each of the next three words (Word, Word + 1, Word + 2, and Word + 3). For the unfronted conditions, disambiguating preposition pairs

TABLE 1
 Experiment 1: Mean Self-Paced Reading Times
 (in Milliseconds) by Condition

	<i>PREP</i>	<i>VERB</i>		
Unfronted/unbounded	486	561		
Unfronted/bounded	509	544		
Fronted/unbounded	466	537		
Fronted/bounded	455	511		
	<i>VERB</i>	<i>VERB+1</i>	<i>VERB+2</i>	<i>VERB+3</i>
Fronted/unbounded	537	549	488	425
Fronted/bounded	511	503	490	420
	<i>PREP</i>	<i>PREP+1</i>	<i>PREP+2</i>	<i>PREP+3</i>
Unfronted/unbounded	486	465	502	445
Unfronted/bounded	509	451	504	456

were compared, and for the fronted conditions, verb pairs were compared. We conducted these analyses on pairs of conditions because the next three words differed between the pairs of conditions and because two or three words were presented together in one pair but not the other.

For the Verb region, 2 (location: fronted vs. unfronted) \times 2 (verb type: bounded vs. unbounded) analyses of variance (ANOVAs) with both factors treated as within-participants and items revealed a marginal effect of fronting, $F_1(1, 55) = 3.97, p = .051$; $F_2(1, 23) = 3.28, p = .083$, suggesting that participants may have taken longer to read the verb when it appeared earlier in the sentence than when it appeared later. No other effects approached significance by either participants or items: verb type, $F_1(1, 55) = 2.62, p = .11$; $F_2 < 1$; interaction, both $F_s < 1$. For the Preposition region, we found an effect of location, $F_1(1, 55) = 4.60, p < .05$; $F_2(1, 23) = 5.27, p < .05$, with participants taking longer to read the preposition when it appeared later in the sentence than when it appeared earlier. No other effects approached significance by either participants or items: verb type, both $F_s < 1$; interaction, $F_1(1, 55) = 2.20, p = .14$; $F_2(1, 23) = 1.14, p = .23$.

No effects approached significance for the unfronted conditions: Preposition, $t(55) = -1.37, p = .18$; $t(23) = -1.12, p = .28$; Preposition + 1, Preposition + 2, Preposition + 3, all $t_s < 1$. For the fronted conditions, t tests revealed that participants spent longer reading the word following the verb (Verb + 1) in the unbounded condition rather than the bounded condition, $t_1(55) = 2.83, p < .01$; $t_2(23) = 2.35, p < .05$. This effect is not predicted by any account and provides no indication that constructing an unbounded interpretation for a bounded verb is difficult. No effects approached significance for the other regions in the fronted conditions: Verb, $t_1(55) = 1.37, p = .18$; $t_2(23) < 1$; Verb + 2, Verb + 3, all $t_s < 1$.

In summary, Experiment 1 revealed no difficulty with the bounded conditions whether fronted or not. The experiment provided some evidence that verbs and prepositions were processed differently depending on where they were in the sentence with the verb showing a tendency toward being processed more straightforwardly when it occurred later in the sentence, and the preposition being processed more straightforwardly when it occurred earlier in the sentence. The former result is more typical of reading (e.g., Haberlandt & Graesser, 1985) with the latter finding perhaps a result of the preposition being the first word in the sentence. Also, one isolated result suggested more difficulty with the unbounded rather than the bounded conditions contrary to any predictions. Therefore, Piñango et al.'s (1999) results may not generalize to normal reading, and people may not commit to a bounded or unbounded interpretation of a sentence during early processing. Because this conclusion largely rests on null results and because it was based on self-paced reading, we decided to replicate the experiment using eye tracking.

EXPERIMENT 2: EYE TRACKING OF PIÑANGO ET AL. (1999)

Experiment 2 was simply a replication of Experiment 1, using eye tracking rather than self-paced reading.

Method

Participants. Forty-eight undergraduates from the University of California at Davis participated in exchange for course credit. All of the participants were native English speakers with normal uncorrected vision.

Stimuli. The stimuli were the same as in Experiment 1 (see 4a–d and Appendix A).

Procedure. A Fourward Technologies Dual-Purkinje Image eye tracker monitored participants' eye movements while they read the test sentences. The tracker has an angular resolution of 10' of arc. The tracker monitored only the right eye's gaze location. A computer screen displayed materials on a monitor 70 cm from participants' eyes. The location of participants' gaze was sampled every millisecond and the computer software recorded the tracker's output to establish the sequence of eye fixations and their start and finish times. Before the experiment, the experimenter seated the participant at the eye tracker and used a bite plate and head rests to minimize head movements. After the tracker was aligned and calibrated, the experiment began. After reading each trial, the participant pressed a key to make the text disappear. After some of the filler sentences, the participant re-

sponded to a comprehension question. Participants received feedback on their responses. All of the participants in the following analyses scored at 90% accuracy or higher on the comprehension questions. Between each trial, a pattern of squares appeared on the computer screen along with a cursor that indicated the participants' current gaze location. If the tracker was out of alignment, the experimenter recalibrated it before proceeding with the next trial.

Analyses. An automatic software procedure pooled short contiguous fixations. Fixations that were less than 80 ms and within one character of another fixation were incorporated into one larger fixation. Fixations of less than 100 ms and not within one character space of another fixation were deleted because, presumably, readers extract little if any information during these short fixations (see Rayner & Pollatsek, 1989). We also excluded fixations of more than 800 ms (e.g., Niswander, Pollatsek, & Rayner, 2000). Because most regions consisted of two or more words, we set the maximum cutoff for first-pass times at 1,800 ms, and at 3,600 ms for total times and regression-path times. Analyses with lower cutoff times did not change the pattern of results.

Results and Discussion

We computed four standard dependent measures. *First-pass time* is the sum of all fixations from the reader's first fixation in the region until the reader fixates outside the region. *First-pass regressions* is the percentage of trials on which a first-pass fixation in the region is immediately followed by a fixation to the left of the region. *Regression-path time* is the sum of all fixations from the reader's first fixation in the region until the reader fixates to the right of the region (i.e., including fixations following a first-pass regression). *Total time* is the sum of all fixations in a region. We analyzed three scoring regions. The *verb* region included the main-clause verb (e.g., *glided* or *hopped*). The *postverb* region included the two words immediately following the verb (*effortlessly under* vs. *effortlessly until*). The *adverbial* region included the adverbial phrase that preceded or followed the main verb (e.g., *until it reached the far end of the garden*). Table 2 presents mean first-pass times, percentage of first-pass regressions, regression-path times, and total times for the three scoring regions.

The data from each scoring region were subjected to separate 2 (verb type: bounded vs. unbounded) \times 2 (location: fronted vs. unfronted) ANOVAs. In the verb region, we found a main effect of Location in the regression-path analysis with participants taking longer when the verb preceded the adverbial (unfronted conditions) than when it followed the adverbial (fronted conditions), $F_1(1, 47) = 4.18, p < .05$; $F_2(1, 23) = 4.20, p = .05$. This effect is hard to interpret because the prior context is longer when the verb followed the adverbial versus when it preceded the adverbial. In addition, the effect did not occur in total time. Indeed no

TABLE 2
 Experiment 2: Mean Reading Times (in Milliseconds)
 and Percentages of First-Pass Regressions

	<i>First Pass</i>	<i>Regressions</i>	<i>Regression Path</i>	<i>Total Time</i>
Verb region				
Unfronted/unbounded	369	6.3	398	445
Unfronted/bounded	340	8.3	397	434
Fronted/unbounded	336	6.3	368	398
Fronted/bounded	328	5.2	361	421
Postverb region				
Unfronted/unbounded	433	14.2	422	542
Unfronted/bounded	423	14.9	538	548
Fronted/unbounded	418	10.1	477	503
Fronted/bounded	419	13.5	520	509
Adverbial phrase				
Unfronted/unbounded	852	9.0	939	954
Unfronted/bounded	892	5.9	946	985
Fronted/unbounded	985	—	985	1,032
Fronted/bounded	972	—	972	1,030
Adverb				
Unfronted/unbounded	271	5.2	321	304
Unfronted/bounded	276	2.8	309	307
Fronted/unbounded	281	—	281	299
Fronted/bounded	279	—	279	312
Postadverb				
Unfronted/unbounded	382	7.6	419	434
Unfronted/bounded	385	7.6	427	449
Fronted/unbounded	359	3.8	380	404
Fronted/bounded	354	4.5	404	417

other main effects or interactions were found in the region. Additionally, no effects emerged in the postverb region.

First-pass time data from the adverb region produced a significant main effect of Location. The adverb region was read slower when it was at the beginning of the sentence than when it followed the main verb (978 vs. 872 ms), $F_1(1, 47) = 7.23$, $p < .01$; $F_2(1, 23) = 45.1$, $p < .0001$. This pattern is not particularly interesting in relation to the processing hypotheses sketched out earlier, and it is consistent with standard order effects in reading (Haberlandt & Graesser, 1985). No other main effects or interactions for the first-pass time, total time, and regression-path time data were significant by both participants and items. In particular, the F values for the interactions were all less than 1.8, except for first-pass time by items, $F_1(1, 47) < 1$; $F_2(1, 23) = 3.08$, $p = .09$. The first-pass regression analyses were computed for the unfronted conditions alone (in the fronted conditions, this was the initial region), $F_1(1, 47) = 3.13$, $p = .083$; $F_2(1, 23) = 1.61$, ns . Notice that the trend is for more regressions to occur in the unbounded than the bounded condition.

To be absolutely sure we had not missed any early effects, we also analyzed two further regions, one consisting of just the *adverb* (e.g., *until*), and the other consisting of the next two words (*postadverb*, e.g., *it reached*). Table 2 gives the means below the dashed line. Because of the design of the experiment, the second word of the postadverb region was not always the same for the fronted and unfronted conditions. However, limiting the postadverb region to just one word would have resulted in many short regions and, thus, many skips. In any case, one should treat analyses of the postadverb region cautiously because they do not always have the same words. For the adverb region, first-pass and total time data showed no effects (all $F_s < 1$). A weak tendency was found toward more regressions in the unbounded condition (5.2%) than the bounded condition (2.8%), again contrary to prediction, $F_1(1, 47) = 3.00, p = .09$; $F_2(1, 23) = 3.48, p = .08$. Regression-path time analyses were also nonsignificant ($ps > .10$). For the postadverb region, weak trends were found on all reading times for effects of Location, with longer reading times for the postadverb region in the unfronted conditions, although none of them reached significance. We found, however, significantly more regressions from this region for the unfronted conditions compared to the fronted conditions (7.6% vs. 4.2%), $F_1(1, 47) = 4.80, p = .03$; $F_2(1, 23) = 5.45, p = .03$. This is presumably because the region is much later in the sentence for the unfronted than the fronted conditions.

Experiment 2 produced a similar pattern of results to Experiment 1. In particular, no evidence was found for difficulty associated with the bounded conditions versus the unbounded conditions. Thus, the results did not replicate Piñango et al.'s (1999) findings. The only reliable effects related to the difficulty of reading regions of the sentence when at the beginning versus the end of the sentence. They therefore suggest that people do not fully commit to an aspectual interpretation during initial processing and hence that they may underspecify for aspect during initial processing. However, Piñango et al.'s (1999) stimuli may not have been sufficiently sensitive to reliably demonstrate such effects. To increase our confidence in this conclusion, we therefore conducted a different test of aspectual processing, using items based on Todorova et al. (2000b).

EXPERIMENT 3: SELF-PACED READING EXPERIMENT BASED ON TODOROVA ET AL. (2000b)

Todorova et al. (2000b) reported two experiments testing aspectual coercion with the stop-making-sense paradigm and found some evidence for difficulty with the adverbial phrase when it required computing the iterative interpretation (and much weaker evidence for similar difficulty at the object noun phrase). Given the lack of effects of aspectual clash in Experiments 1 and 2 herein and the concern that the stop-making-sense task may not reflect normal reading processes, we decided to

conduct a self-paced reading study (with no sensibility task during or after the sentence) based on a combination of conditions from their two experiments. The strongest test of whether an aspectual clash causes difficulty involves items with frequency modifiers (e.g., *every year*) because they avoid the potential ambiguity introduced by durative modifiers (e.g., *for many years*), and items with neutral (nondurative) adverbs (e.g., *last year*) because these serve as an appropriate control. This contrasts with Todorova et al. (2000b), Experiment 2, which contrasted frequency and durative modifiers. Additionally, we modified their items somewhat, both to enhance their plausibility and to introduce a spillover region following the adverbial.

An example item is (5a–d) with slashes defining presentation regions:

- (5a) Howard sent / a large check / to his daughter / every year / but as / usual, she refused / to accept his money. Singular object–frequency adverb (SF).
- (5b) Howard sent / large checks / to his daughter / every year / but as / usual, she refused / to accept his money. Plural object–frequency adverb (PF).
- (5c) Howard sent / a large check / to his daughter / last year / but as / usual, she refused / to accept his money. Singular object–neutral adverb (SN).
- (5d) Howard sent / large checks / to his daughter / last year / but as / usual, she refused / to accept his money. Plural object–neutral adverb (PN).

In (5a and c), *sent a large check* is bounded. This is compatible with *last year* in (5c), but not with *every year* in (5a). If readers have adopted this interpretation in (5a), they will then need to reinterpret it iteratively. In (5b and d), *sent large checks* may be unbounded (assuming the checks are sent successively). If the aspectual clash the adverb introduces causes difficulty, then participants should experience difficulty with (5a) versus (5b) at or soon after the adverbial *every year*, but should not experience comparable difficulty at or soon after the adverbial *last year* in (5c–d). (They may also experience difficulty at or soon after reading *large checks* in (5b and d) versus *a large check* in (5a and c), but note that the object itself differs between the pairs of conditions.) If on the other hand readers do not commit fully to an aspectual interpretation of *sent a large check*, then they should experience no difficulty with (5a) after encountering *every year*.

Method

Participants. Sixty native English speaking students from New York University were paid to participate in the experiment. All had normal or corrected-to-normal vision.

Items. These were based on Todorova et al. (2000a, 2000b) as described earlier (see Appendix B for sample items and questions). They were divided into four lists so that each participant saw nine items in each condition and one version of

each item. The 36 critical sentences were mixed with 120 sentences of various types. The sentences were presented randomly.

Procedure. The procedure was the same as in Experiment 1, except that instead of (mainly) single word regions, we used larger regions (mainly) corresponding to individual phrases (see Appendix B). We also did not use the cumulative accuracy percentage feedback of Experiment 1 but rather calculated accuracy at the end of the experiment. Accuracy for the comprehension questions was 96.25%.

Results and Discussion

We set the cutoff points at 100 ms and 4,000 ms, which eliminated less than 1% of the data (analyses with different cutoff points produced comparable results). The upper cutoff was higher than in Experiment 1 because Experiment 3 included longer regions than Experiment 1. We report analyses on 5 regions: the object (*a large check, large checks*), the prepositional phrase (*to his daughter*), the temporal adverbial (*every year, last year*), a first spillover region (*but as*) and a second spillover region (*usual, she refused*). Table 3 presents the means for these regions. For each region, we subjected the data to separate Object (singular vs. plural) \times Adverb (durative vs. neutral) ANOVAs.

We found no significant effects on any region, including the Verb region (all $ps > .18$), the Object region (all $ps > .12$), the Prepositional Phrase region (all $ps > .19$), the Temporal Adverbial phrase (all $ps > .24$), the first spillover region (all $ps > .12$), and the second spillover region (all $ps > .24$). In particular, the interaction in the critical Temporal Adverbial phrase region did not approach significance, $F_1(1, 59) = 1.07, p = .31$; $F_2(1, 35) < 1$. To be absolutely certain that we had not missed any effects, we conducted means comparisons between the singular and plural object conditions separately for the durative adverbial conditions and for the neutral adverbial conditions, but none of the comparisons approached significance (all $ps > .29$).

These results therefore provide further reasons to doubt that aspectual clashes cause processing difficulty during self-paced reading, at least for items similar to

TABLE 3
Self-Paced Reading Times for Experiment 3

	<i>Object</i>	<i>Prep</i>	<i>Adverb</i>	<i>Spill 1</i>	<i>Spill 2</i>
S-F	588	734	579	415	494
P-F	602	759	598	404	493
S-N	618	772	578	408	484
P-N	615	766	569	416	482

Note. S = single object; P = plural object; F = frequency adverb; N = neutral adverb.

those Todorova et al. (2000a, 2000b) used. To determine whether we can be fully confident about this conclusion, we conducted an eye-tracking experiment using the same items.

EXPERIMENT 4: EYE-TRACKING EXPERIMENT, BASED ON TODOROVA ET AL. (2000b)

Method

Participants. Forty native English speaking students from New York University were paid to participate in the experiment. All had normal or corrected-to-normal vision.

Materials. We used the same 36 item quartets from Experiment 3, divided into four lists, and added 176 filler sentences of different types.

Procedure. Participants were run individually using a SensoriMotor Instruments EyeLink I head-mounted eye-tracker apparatus and presentation software. The eye cameras recorded eye movements and fixations every 4 ms. Viewing was binocular, but only data from the eye that was calibrated best was used in the analyses. Screen resolution was set at $1,600 \times 1,200$ pixels and sentences were displayed in Courier New fixed font with each letter being 18 pixels wide and 33 pixels high and a maximum of 80 characters per line. Participants were seated 28 in. from the display monitor, and a chin rest reduced head movements. With this setup, 1° of visual angle subtended 2.7 characters. Calibration was performed at the beginning of the experiment and whenever the experimenter felt necessary. At the start of each trial, participants looked at calibration point, which coincided with the first letter of the upcoming sentence, and a drift correction was performed. The entire experiment lasted about 40 min.

Analyses. These were the same as Experiment 2. We report analyses on four regions: the object, the prepositional phrase, the temporal adverbial, and a spillover region (defined as the next word if five letters or more, otherwise the next two words). The same measures as in Experiment 2 are discussed: first-pass time, percentage of first-pass regressions, regression-path time, and total time. The reported means and ANOVAs are based on reading times excluding zero fixations.

Results and Discussion

Prior to all analyses, sentences with major track losses due to head movements or blinks and sentences for which two consecutive regions were skipped were ex-

cluded from the analyses (2.6% of the data). For each measure and each region, the data were subjected to separate Object (singular vs. plural) \times Adverb (frequency vs. neutral adverbial) ANOVAs, treating participants (F_1) and items (F_2) as random effects. Table 4 presents the (participant) means. We discuss each scoring region separately.

For the direct object region, we observed an effect of Object (marginal by participants) with 20 ms longer reading times when the object was used singularly, $F_1(1, 39) = 3.39, p = .07$; $F_2(1, 35) = 4.33, p < .05$. Because this region was slightly longer for the singular objects, this difference might represent a mere length effect. No significant effects appeared on the first-pass regression measure (all $F_s < 1$), the regression-path time measure (all $F_s < 1$), and the total time measure (all $p_s > .17$). For the prepositional phrase region, an effect of Object was observed in first-pass time, but on this occasion the conditions containing the plural object were read 23 ms slower, $F_1(1, 39) = 5.49, p = .02$; $F_2(1, 35) = 5.62, p = .02$. At the same time, the plural object conditions triggered marginally fewer first-pass regressions compared to the singular object conditions: 8.8% vs. 10.8%, $F_1(1, 39) = 2.10, p = .16$; $F_2(1, 35) = 3.17, p = .08$. No effects emerged on the regression-path time measure (all $F_s < 1$) or the total time measure (all $F_s < 1$). For the temporal ad-

TABLE 4
Reading Times and Percentages of First-Pass Regressions
for Experiment 4

	<i>First Pass</i>	<i>Regressions</i>	<i>Regression Path</i>	<i>Total Time</i>
Object region (<i>a large check/large checks</i>)				
S-F	337	12.9	427	464
P-F	313	13.8	416	428
S-N	336	14.1	416	468
P-N	320	15.7	426	453
Prepositional phrase (<i>to his daughter</i>)				
S-F	450	11.3	552	621
P-F	485	8.7	570	615
S-N	471	10.4	567	632
P-N	481	8.0	558	616
Temporal adverbial (<i>every year/last year</i>)				
S-F	362	5.4	552	435
P-F	354	5.3	570	437
S-N	339	6.5	567	413
P-N	356	7.6	558	440
Spillover (<i>but as</i>)				
S-F	279	4.0	316	331
P-F	283	4.5	319	333
S-N	270	1.9	298	317
P-N	278	2.8	307	330

Note. S = single object; P = plural object; F = frequency adverb; N = neutral adverb.

verbal region, no significant effects were found for the first-pass time measure (all $ps > .08$) and, crucially, means comparisons showed no statistical difference between the singular and the plural object conditions with a frequency adverb (S-F vs. P-F): both $ts < 1$. No significant effects were found for the first-pass regressions measure for this region (all $ps > .18$). The regression-path time measure for this region showed a marginally significant interaction, $F_1(1, 39) = 2.68, p = .11$; $F_2(1, 35) = 5.67, p = .02$, although the pattern differed from the one the immediate commitment account predicted: No differences occurred between the conditions with a frequency adverb ($ts < 1$), although there was a suggestion that the plural object, neutral adverb condition took longer to process than the singular object, neutral adverb condition, $t(39) = 1.68, p = .10$; $t(35) = 2.29, p = .03$. No effect emerged for this region on the total time measure (all $ps > .18$). For the spillover region, no effects were found for first-pass time (all $F_s < 1$). There was some suggestion that the singular object conditions triggered more first-pass regressions than the plural object conditions, $F_1(1, 39) = 5.46, p = .03$; $F_2(1, 35) = 2.36, p = .13$, although given the very low number of regressions from this region, one should be careful interpreting this effect. No effects occurred for the regression-path time measure (all $ps > .18$) and the total time measure (all $F_s > 1$).

To summarize, the results from this experiment mirror the self-paced reading results (Experiment 3) and, again, fail to show any difficulty for an aspectual clash as assumed by the immediate commitment hypothesis. There was some suggestion that the preferred telic construction (singular object) was *harder* than the dispreferred one, but this could be related to length as well (or conceptual oddness).

GENERAL DISCUSSION

We have reported four experiments that address the question of whether readers experience difficulty with what are argued to be aspectual mismatches between a verb phrase and a durative adverbial phrase. In contrast to earlier studies by Piñango et al. (1999) and Todorova et al. (2000a, 2000b), we found no evidence that this type of semantic mismatch engendered any detectable differences in conventional reading tasks.

Why are our data so different from these other studies? Our results are essentially null results, and so a concern always exists that our failure to reject the null hypothesis might simply represent a Type II error. However, that four separate experiments using two techniques and four sets of participants from two educational institutes all gave rise to Type II errors is very unlikely. Each experiment used a large number of participants and items, and employed techniques that are known to be highly sensitive to effects of different types of language manipulations (e.g., Rayner, 1998). Moreover, as part of other ongoing research, all four experiments

included constructions involving another type of semantic coercion, namely complement coercion such as (6a):

- (6a) The author began the book. ...
- (6b) The author read the book. ...

Sentences such as (6a) involve a type of coercion because the verb *began* requires an event complement, but here the object *the book* refers to an entity rather than an event. To interpret such an expression, readers have to coerce the complement into an event, for example *reading the book* or *writing the book* (Jackendoff, 1997; Pustejovsky, 1995). In previous studies (e.g., McElree, Frisson, et al., 2006; McElree, Pykkänen, et al., 2006; McElree et al., 2001; Pickering et al., 2005; Traxler et al., 2002, 2005), we have found reliable differences between conditions such as (6a) and control expressions such as (6b). The same effects were found in all four studies reported here. This provides direct evidence that the experiments were sufficiently sensitive to detect difficult types of interpretive operations and that the null results for expressions with aspectual coercion do not reflect aberrant participant populations or faulty measurement procedures.

Our results are most consistent with the notion that comprehenders routinely underspecify aspectual properties of an interpretation during normal reading. Effects of aspectual coercion appear to engender measurable differences only when reading is paired with a concurrent task, such as a lexical-decision task (Piñango et al., 1999) or a stop-making-sense task (Todorova et al., 2000a, 2000b). We believe that these secondary tasks induce comprehenders to commit to properties of an interpretation that they would otherwise leave unspecified. In Todorova et al.'s task, participants both read and make judgments about the plausibility of what they are reading. Although participants might normally underspecify aspect while reading, the stop-making-sense judgment may induce them to resolve aspect immediately. Hence, this task is more likely to lead to full commitment—or at least commitment much earlier than our tasks. Similarly, Piñango et al.'s task requires people to perform a second task while listening to sentences, and such additional processing might likewise trigger people into earlier commitment than otherwise because their attention is drawn to the part of the sentence involving aspectual coercion or because they realize that they would not be likely to have the processing resources available later in the sentence to delay the act of coercion.

Although difficulty might have conceivably emerged if our comprehension questions had uniformly required deep processing or focused on the aspectual coercion itself, one should note that Piñango et al. (1999) and Todorova et al. (2000a, 2000b) did not use such “deep” questions either. Additionally, such deep questions do not regularly occur when readers encounter connected discourse. Our results therefore provide no reason to believe that readers need to commit fully to the telicity of events, although such commitment can possibly occur under some read-

ing conditions, for example, when such a decision is critically important for providing coherence to a connected text or when it is morphologically marked.

Whether task demands are solely responsible for the differences between our results and those of Piñango et al. (1999) and Todorova et al. (2000b) is a matter of debate. However, importantly, our results with expressions involving aspectual coercion contrast dramatically with our results with expressions involving complement coercion. The latter have consistently shown reliable effects in self-paced reading and eye-tracking tasks (McElree, Frisson, et al., 2006; McElree et al., 2001; Pickering et al., 2005; Traxler et al., 2002, 2005), as well as with other methods and measures (McElree, Pykkänen, et al., 2006; Pykkänen, Llinas, & McElree, 2004). Minimally, the lack of reading time effects for aspectual coercion demonstrates that this type of semantic mismatch is simpler or easier for the processing system to deal with than the semantic mismatches found in complement coercions such as *began the book*. However, we believe that this contrast may also point to a more fundamental difference in basic compositional operations.

Complete exposition of the differences between aspectual and complement coercion is beyond the scope of this article. However, Pykkänen and McElree (in press) situate both types of expressions within recent compositional approaches to semantics. They argue that interpreting aspectual mismatches such as *hopped until* or *sent large checks for years* does not require shifting the basic ontological types of the constituents in the expression, but only fixing of various contextual parameters on the basis of world knowledge (how long the event lasted, whether the event was repetitive or occurred only once, etc.). The latter may allow for degrees of underspecification because fixing all parameters in every situation may not be necessary.

In contrast, Pykkänen and McElree (in press) suggested that interpreting an expression such as *began the book* requires comprehenders to shift the ontological type of the complement, from the default type ENTITY denoted by the noun to the type EVENT required by the verb. Without doing so, a basic interpretation of the verb phrase is blocked because the complement cannot be unified with the semantic requirements of the eventive verb. This requirement disallows radical underspecification because it forces comprehenders to resolve immediately the semantic mismatch to derive a coherent interpretation. Our work on complement coercion suggests that it engenders a consistent processing cost because comprehenders need to engage immediately in additional online operations, such as the generation of unexpressed semantic structure, to construct an eventive interpretation of the complement (see McElree, Frisson, et al., 2006; Traxler et al., 2002, 2005). Although comprehenders may underspecify components of the event involving the complement, composition mandates the construction of *some* type of eventive representation of the complement. Comprehenders may have the option of not committing to whether *began the book* means, for example, reading the book or writing the book; however, they can not forgo constructing a skeleton event rep-

resentation, one that might be paraphrased as “began doing something with or to the book.” No comparable act of type-shifting is needed to interpret aspectual coercions such as *hopped until*, nor do they seem to involve the construction of additional semantic structure.

We believe that differences in the degree to which aspectual and complement coercion allow for underspecification highlight fundamental differences in semantic interpretations. If this interpretation is correct, it illustrates how the investigations of limits on incrementality can serve as a useful means for exploring different facets of semantic processing.

ACKNOWLEDGMENTS

This research was supported by National Science Foundation Grant BCS-0236732 awarded to Brian McElree. We acknowledge Marina Todorova for making her items available to us.

REFERENCES

- Barton, S. B., & Sanford, A. J. (1993). A case-study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, *21*, 477–487.
- Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from *wh*-questions. *Journal of Memory and Language*, *34*, 774–806.
- Erickson, T. A., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, *20*, 540–552.
- Fodor, J., Bever, T., & Garrett, M. (1974). *The psychology of language*. New York: McGraw-Hill.
- Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage*. New York: Houghton Mifflin.
- Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language*, *29*, 181–200.
- Frazier, L., Pacht, J., & Rayner, K. (1999). Taking on semantic commitments, II: Collective vs. distributive readings. *Cognition*, *70*, 87–104.
- Frisson, S., & Pickering, M. J. (1999). The processing of metonymy: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1366–1383.
- Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General*, *114*, 357–374.
- Jackendoff, R. (1997). *The architecture of language faculty*. Cambridge, MA: MIT Press.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329–354.
- Magliano, J. P., Graesser, A. C., Eymard, L. A., Haberlandt, K., & Gholson, B. (1993). Locus of interpretative and inference processes during text comprehension: A comparison of gaze durations and word reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 704–709.

- Marslen-Wilson, W. D. (1973). Linguistic structure and speech shadowing at very short latencies, *Nature*, 244, 522–523.
- McElree, B., Frisson, S., & Pickering, M. J. (2006). Deferred interpretations: Why starting Dickens is taxing but reading Dickens isn't. *Cognitive Science*, 30, 181–192.
- McElree, B., Pylkkänen, L., Pickering, M. J., & Traxler, M. J. (2006). The time course of enriched composition. *Psychonomic Bulletin & Review*, 13, 53–59.
- McElree, B., Traxler, M. J., Pickering, M. J., Seely, R., & Jackendoff, R. (2001). Reading time evidence for enriched composition. *Cognition*, 78, B17–B25.
- Niswander, E., Pollatsek, A., & Rayner, K. (2000). The processing of derived and suffixed words during reading. *Language and Cognitive Processes*, 15, 389–420.
- Pickering, M. J., & Frisson, S. (2001). Processing ambiguous verbs: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 556–573.
- Pickering, M. J., Traxler, M. J., & McElree, B. (2005). The difficulty of coercion: A response to de Almeida. *Brain and Language*, 93, 1–9.
- Piñango, M., Zurif, E., & Jackendoff, R. (1999). Real-time processing implications of enriched composition at the syntax-semantics interface. *Journal of Psycholinguistic Research*, 28, 395–414.
- Proctor, A. S., Dickey, M. W., & Rips, L. (2004, August). *The time-course and cost of telicity inferences*. Proceeding of the 26th Annual Meeting of the Cognitive Science Society, Chicago.
- Pustejovsky, J. (1995). *The generative lexicon*. Cambridge, MA: MIT Press.
- Pylkkänen, L., Llinas, R., & McElree, B. (2004, August). *Distinct effects of semantic plausibility and semantic composition in MEG*. In E. Halgren, S. Ahlfors, M. Hämäläinen, & D. Cohen (Chairs), *Biomag 2004: Proceedings of the 14th International Conference on Biomagnetism*, Biomag, Boston.
- Pylkkänen, L., & McElree, B. (in press). The syntax-semantics interface: On-line composition of sentence meaning. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed.). New York: Elsevier.
- Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Todorova, M., Straub, K., Badecker, W., & Frank, R. (2000a, August). *Aspectual coercion and the on-line computation of sentential aspect*. Proceedings of the 22nd Annual Conference of the Cognitive Science Society (pp. 3–8).
- Todorova, M., Straub, K., Badecker, W., & Frank, R. (2000b, August). *Processing correlates of aspectual computation*. Presented to the Workshop on Events and Paths, ESSLLI XII, Birmingham, England.
- Traxler, M. J., McElree, B., Williams, R. S., & Pickering, M. J. (2005). Context effects in coercion: Evidence from eye-movements. *Journal of Memory and Language*, 53, 1–25.
- Traxler, M. J., Pickering, M. J., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language*, 4, 530–547.
- Wason, P. C. (1959). The processing of positive and negative information. *Quarterly Journal of Experimental Psychology*, 11, 92–107.

APPENDIX A

Sample items with comprehension questions for Experiments 1 and 2. For each item, the fronted–unbounded condition appears first (a), followed by the fronted–bounded condition (b), the unfronted–unbounded condition (c), and the unfronted–bounded condition (d).

- (1a) Until well into the evening, the boy screamed loudly just to annoy his mother. He was always a very noisy and irritable child.
- (1b) Until well into the evening, the boy belched loudly just to annoy his mother. He was always a very noisy and irritable child.
- (1c) The boy screamed loudly until well into the evening at which point he was satisfied. He was always a very noisy and irritable child.
- (1d) The boy belched loudly until well into the evening at which point he was satisfied. He was always a very noisy and irritable child.

Question: *Was the boy irritable?* (Yes)

- (2a) For an hour, the tiger slept near the bushes after eating the biggest meal of his life. He enjoyed every bit of it.
- (2b) For an hour, the tiger jumped near the bushes after eating the biggest meal of his life. He enjoyed every bit of it.
- (2c) The tiger slept near the bushes for an hour after eating the biggest meal of his life. He enjoyed every bit of it.
- (2d) The tiger jumped near the bushes for an hour after eating the biggest meal of his life. He enjoyed every bit of it.

Question: *Did a giraffe sleep/jump near the bushes?* (No)

- (3a) Until the teacher told her she should eat, the little girl dogpaddled awkwardly in the pool. Although she was hungry, she continued to swim quietly.
- (3b) Until the teacher told her she should eat, the little girl dived gracefully in the pool. Although she was hungry, she continued to swim quietly.
- (3c) The little girl dogpaddled awkwardly until the teacher told her she should eat. Although she was hungry, she continued to swim quietly.
- (3d) The little girl dived gracefully until the teacher told her she should eat. Although she was hungry, she continued to swim quietly.

Question: *Was the girl hungry?* (Yes)

- (4a) For hours, the man meandered aimlessly in the park after receiving the good news. His wife had just bore a baby boy.
- (4b) For hours, the man sneezed violently in the park after receiving the good news. His wife had just bore a baby boy.
- (4c) The man meandered aimlessly for hours in the park after receiving the good news. His wife had just bore a baby boy.
- (4d) The man sneezed violently for hours in the park after receiving the good news. His wife had just bore a baby boy.

Question: *Did his wife bear a baby girl?* (No)

APPENDIX B

Sample items with comprehension questions for Experiments 3 and 4. For each item, the singular object–frequency adverb condition appears first (a), followed by the plural object–frequency adverb condition (b). The (c and d) versions are control conditions, with (c) being the singular object–neutral adverb and (d) being the plural object–neutral adverb condition.

- (1a) Ronnie broke a hockey stick on the ice every season but this did not stop him to continue to practice aggressively.
- (1b) Ronnie broke hockey sticks on the ice every season but this did not stop him to continue to practice aggressively.
- (1c) Ronnie broke a hockey stick on the ice last season but this did not stop him to continue to practice aggressively.
- (1d) Ronnie broke hockey sticks on the ice last season but this did not stop him to continue to practice aggressively.

Question: *Did Ronnie break a baseball bat?* (No)

- (2a) John threw a football across the field every afternoon but the range and accuracy of his passes remained unstable.
- (2b) John threw footballs across the field every afternoon but the range and accuracy of his passes remained unstable.
- (2c) John threw a football across the field before the game but the range and accuracy of his passes remained unstable.
- (2d) John threw footballs across the field before the game but the range and accuracy of his passes remained unstable.

Question: *Was John's accuracy unstable?* (Yes)

- (3a) Derek hit a home run in the practice field every week but then he struck out during the important game.
- (3b) Derek hit home runs in the practice field every week but then he struck out during the important game.
- (3c) Derek hit a home run in the practice field last week but then he struck out during the important game.
- (3d) Derek hit home runs in the practice field last week but then he struck out during the important game.

Question: *Did Derek hit a home run during the important game?* (No)

- (4a) Howard sent a large check to his daughter every year but as usual, she refused to accept his money.
- (4b) Howard sent large checks to his daughter every year but as usual, she refused to accept his money.
- (4c) Howard sent a large check to his daughter last year but as usual, she refused to accept his money.
- (4d) Howard sent large checks to his daughter last year but as usual, she refused to accept his money.

Question: *Was (Were) the check(s) large?* (Yes)