

Eye movement evidence that readers maintain and act on uncertainty about past linguistic input

Roger Levy^{a,1}, Klinton Bicknell^a, Tim Slattery^b, and Keith Rayner^b

Departments of ^aLinguistics and ^bPsychology, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093

Edited by Edward E. Smith, Columbia University, New York, NY, and approved October 20, 2009 (received for review July 14, 2009)

In prevailing approaches to human sentence comprehension, the outcome of the word recognition process is assumed to be a categorical representation with no residual uncertainty. Yet perception is inevitably uncertain, and a system making optimal use of available information might retain this uncertainty and interactively recruit grammatical analysis and subsequent perceptual input to help resolve it. To test for the possibility of such an interaction, we tracked readers' eye movements as they read sentences constructed to vary in (i) whether an early word had near neighbors of a different grammatical category, and (ii) how strongly another word further downstream cohered grammatically with these potential near neighbors. Eye movements indicated that readers maintain uncertain beliefs about previously read word identities, revise these beliefs on the basis of relative grammatical consistency with subsequent input, and use these changing beliefs to guide saccadic behavior in ways consistent with principles of rational probabilistic inference.

psycholinguistics | language comprehension | probabilistic models of cognition

A critical part of understanding a sentence is identifying the words that comprise it. It is well understood that humans achieve this goal through a combination of top-down (contextual) and bottom-up (perceptual) cues (1, 2). In spoken word recognition, comprehenders' conclusions about a given word's identity can be affected by phonetic and lexical cues from adjacent words (3, 4), and by subsequent semantic context when the speech stream is manipulated to make bottom-up evidence ambiguous (5, 6). It has also recently become clear that even relatively unambiguous bottom-up evidence regarding spoken word identity can be overridden by subsequent evidence within the same word (7, 8). It is unknown, however, whether word-identity uncertainty is maintained as a matter of course after perceptual cues have been encountered in full upon hearing or first reading a word in a sentence. It could be that, under normal circumstances, a categorical commitment is made to a unique candidate word, which is then integrated into an incremental sentence representation. Such a strategy could be efficient if considerable effort is required to construct and maintain syntactic/semantic relationships for multiple candidate words at a given position within the sentence (9). This possibility also coincides with the treatment of grammatical analysis in most formal models of language comprehension as a process that takes a word sequence as its input (10–12), and with the treatment in leading models of eye-movement control in reading (13, 14). On the other hand, it is also possible that the perceptual input from a given word gives rise to multiple candidate word representations entering the stream of syntactic/semantic processing. This strategy would have the advantage of greater robustness, permitting identification and correction of perceptual noise or speaker error early in a sentence on the basis of additional information obtained late in a sentence (15). In addition, maintaining and continually reassessing uncertainty about word identity could be useful in guiding behavior during real-time comprehension. In reading, eye movements could be directed to previous parts of a sentence where word-identity

uncertainty is high; in spoken language comprehension, allocation of attentional resources could be updated on the basis of word uncertainty, or the listener could prompt the speaker for clarification about parts of a message that were unclear. Whether uncertainty about word identity is maintained and revised, and if so, according to what principles, is thus a question of fundamental interest in the study of language comprehension.

A Sharp Test for Word-Level Uncertainty

In the past, providing a sharp test of whether word-identity uncertainty is maintained has proven elusive because existing results could be interpreted as involving categorical word misidentification. For example, the word *flour* is reread more often in sentence 1a than in sentence 1b or than the word *wheat* in sentence 1c (16):

- 1a. He swept the flour that he spilled.
- 1b. The baker needed more flour for the special bread.
- 1c. He swept the wheat that he spilled.

Crucially, the word *flour* has an orthographically similar “near-neighbor,” *floor*, that is semantically incompatible with the continuations of sentences 1a or 1c. The sentence-initial context in sentence 1a is more strongly predictive of *flour* than of *floor*, so it is possible that the distinctive rereading behavior is due to the comprehender initially maintaining some uncertainty as to whether the critical word was *flour* or *floor*, and subsequently returning to this word to confirm that it is indeed *flour*. This same result could, however, be obtained under a theory in which uncertainty about word identity is not maintained after initial word recognition, but there is an occasional categorical misidentification of *flour* as *floor* in sentence 1a upon first reading, and the incongruous downstream word *spilled* triggers a return of the eyes to the initial locus of semantic anomaly.

In the present work, we provide a sharper test of whether word-identity uncertainty is maintained by using sentences in which categorical misidentification should in principle facilitate, rather than hinder, downstream processing. We turn to sentences with a combination of grammatical constructions that has attracted considerable recent attention for the theoretically problematic comprehension behavior it elicits. Despite near meaning equivalence, sentence 2a is considerably more difficult to read and comprehend than sentences 2b and 2c (17–19):

- 2a. The coach smiled at the player tossed the frisbee.
- 2b. The coach smiled at the player thrown the frisbee.
- 2c. The coach smiled at the player who was tossed the frisbee.

Author contributions: R.L., K.B., T.S., and K.R. designed research; R.L., K.B., and T.S. performed research; R.L., T.S., and K.R. contributed new reagents/analytic tools; R.L., K.B., T.S., and K.R. analyzed data; and R.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. E-mail: rlevy@ling.ucsd.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0907664106/DCSupplemental.

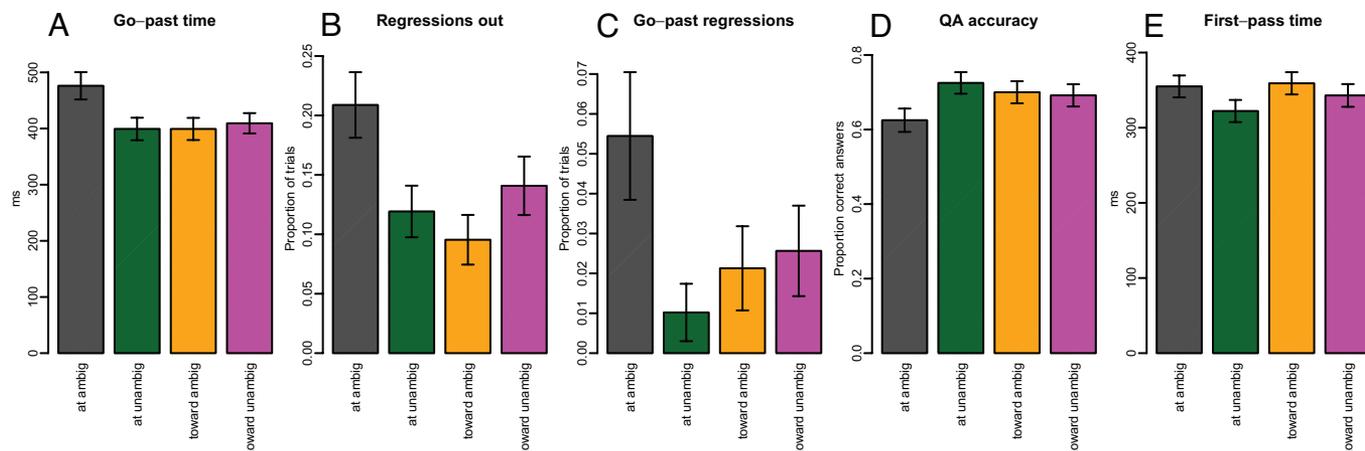


Fig. 2. Means and standard errors of measures of processing difficulty associated with the critical word [e.g., *tossed* in sentences 2a and 3a); *thrown* in sentences 2b and 3b)] and overall sentence comprehension. (A) Proportion of trials with first-pass regression from critical word. (B) Go-past time from first fixation on critical word to first fixation beyond it. (C) Proportion of trials with fixation on earlier preposition (*at/toward*) during go-past reading of critical word. (D) Accuracy in comprehension-question answering. (E) First-pass time on critical word. In A–C, interactions between preposition and critical-word ambiguity are significant (all ANOVA $P < 0.05$); in D, the interaction is $P = 0.087$. In E, main effect of critical-word ambiguity is significant (ANOVA $P < 0.05$ by participants, $P < 0.1$ by items).

variants of these sentences in which the word *at* is substituted by the word *toward*, as in sentences 3a and 3b:

- 3a. The coach smiled toward the player tossed the frisbee.
3b. The coach smiled toward the player thrown the frisbee.

Because *toward* does not have near-neighbor words under which *the player* would be a grammatical subject, the shift in belief about preceding context upon reading the verb *tossed*—but not *thrown*—is considerably reduced (Fig. 1, orange and magenta lines). In behavioral correlates of processing difficulty at this critical verb in the sentence, we should therefore see an interaction between the word used in the preceding context (*at* versus *toward*) and the superficial part-of-speech ambiguity of the word in this position (*tossed*, ambiguous; *thrown*, unambiguous).

Experiment

Our experiment tested the prediction described above of an interaction between near-neighborhood of the preceding context and superficial ambiguity of the critical word on measures of processing difficulty in sentence reading. We constructed 24 sets of sentences on the pattern of sentences 2a and b and 3a and b; the four versions of each sentence differed only in the immediately postverbal preposition (*at* versus *toward*) and whether the critical participial verb had the same surface form as the verb's simple-past form. We tracked the eye movements of 40 native English speakers as they read these sentences on a computer screen and answered a yes/no comprehension question about each sentence.

Results. Following standard practice in the analysis of reading using eye tracking, we divided each sentence into seven regions as shown in example 4 and analyzed seven measures of online processing difficulty widely used in the sentence-processing literature (33).

4. Subj MV Prep Obj
/The coach/ smiled/ {at,toward}/ the player/
Critical Spill Final
{tossed,thrown}/ a frisbee/ by the opposing team./

We report here results pertinent to our hypothesis of an interaction between preposition in preceding context (*at* versus *toward*) and superficial ambiguity of the critical verb (ambiguous versus unambiguous). (Complete results are given in *SI Appen-*

dix.) We observed precisely the predicted interaction in three measures of incremental processing difficulty: (i) the frequency with which readers' eye movements immediately after first reading the critical verb were regressive (as opposed to progressive; significant both in by-participant and by-item ANOVAs at $P < 0.05$); (ii) the time spent reading between the first fixation on the critical verb and the first fixation beyond it (significant both in by-participant and by-item ANOVAs at $P < 0.05$); and (iii) the frequency with which readers fixated on the preposition (*at* or *toward*) after first fixating on the critical verb and before fixating beyond it ($P = 0.087$ in a mixed-effects logit model). By all of these measures, readers experienced the most difficulty in the condition where the critical verb is superficially part-of-speech ambiguous and near-neighbor substitutions in the preceding context would license the incorrect part of speech (Fig. 2A–C). Additionally, readers were least likely (ANOVA interactions $P < 0.05$) to correctly answer comprehension questions about these sentences in this condition (Fig. 2D).[†] Finally, we found that readers spent longer overall during first-pass reading of the critical word in the ambiguous conditions (Fig. 2E; ANOVA main effect $P < 0.05$ by participants, $P < 0.1$ by items).[‡]

Discussion These results provide evidence that readers maintain uncertainty about the identities of previously read words and that grammatical relationships that would hold between a word at a given position and the rest of the sentence are entertained for multiple candidate words. As measured by frequency and duration of regressive eye movements, comprehension is most effortful in the *at*+ambiguous condition, precisely the condition where the discrepancy between the likelihood of the critical word given the true preceding context versus near-neighbor contexts most strongly favors the near-neighbor contexts. Unlike previous

[†]Some of our comprehension questions specifically queried the relationship of the critical verb with the preceding noun phrase [e.g., *Did the player (toss/throw) the frisbee?*], and those questions were answered least accurately in the *at*+ambiguous condition, though the three-way interaction was not significant.

[‡]To address possible concerns regarding whether the observed patterns are driven by trials in which participants did not fixate on the preposition during first-pass reading or by a mismatch in main-clause plausibility between *at* and *toward* conditions, we also conducted analyses of the subset of trials in which the preposition was fixated during first-pass reading and of a subset of items for which mean plausibility was matched across *at* and *toward* conditions on the basis of a separate norming study. In both of these analyses, the qualitative patterns were the same as for the complete data, and in some cases they were more highly significant (see *SI Appendix* for complete analysis).

results, this result cannot be explained by occasional categorical word misidentification: misreading *and* or *as* for *at* should make the critical region easier when it is ambiguous (*tossed*) than when it is unambiguous (*thrown*), because the misreading would make a finite-verb analysis available.

General Discussion

Our results suggest that readers, as a matter of course, maintain uncertain beliefs about the identities of previously read words, revise these beliefs as a function of grammatical coherence with subsequent linguistic input, and act rapidly on changes in these beliefs through eye movements directed toward the loci of changes in uncertainty. Any theory of language comprehension in which representations of preceding context involve veridical word representations lacking in uncertainty will have difficulty accommodating these findings. These findings also go beyond previous results demonstrating downstream effects of context in spoken word recognition (3–6) in that our materials are visual rather than auditory, were not manipulated so as to create unusual perceptual ambiguity, and require the simultaneous pursuit of multiple grammatical analyses of the sentence to obtain the belief-revision behavior we observed.

Finally, our results are consistent with a picture in which key elements of language comprehension can be viewed as instances of rational behavior through probabilistic inference (34–36): perceptual input obtained from eye movements is recruited jointly with grammatical knowledge to produce evidential inferences about linguistic form and structure, which in turn play a role in guiding subsequent eye movements. Further research will be necessary to more fully establish the full range of uncertainty that comprehenders hold out as to word identities—whether they consider only substitutions among closed-class words (which would be consistent with the present experimental results) or also entertain substitutions among open-class words, as well as the omission, insertion, and even position swapping of linguistic material within a sentence. Nevertheless, the view implied here of language comprehension as jointly recruiting perceptual input and grammatical knowledge has potentially broad consequences for the study of language and cognition. The view simultaneously unites and underscores the differences between research in spoken language understanding and reading, raising new questions about the relative weighting of sensory input and prior knowledge, and how noise (environmental, neural, and speaker error driven) is accounted for, in the two modalities. Answers to these questions would be of both fundamental value in cognitive science and potential benefit in furthering our understanding of disorders and age-related changes in both reading and spoken language comprehension.

Methods

Materials and Procedure. We used 24 experimental items in the study and constructed four stimulus lists, rotating items among these four conditions in

a Latin Square; complete experimental materials can be found in *SI Appendix*. These 24 experimental stimuli were interleaved with 36 fillers. Order of presentation was randomized differently for each participant, subject to the constraint that no two experimental items appeared consecutively. Forty native-English speaking undergraduate students at the University of California at San Diego participated in the experiment. All had normal vision or corrected to normal vision and were naive as to the purpose of the experiment. Participants read each sentence while their eye movements were monitored by an SR EyeLink 2000 eye tracker, obtaining one eye-position sample every 1/2 ms with a spatial resolution of 0.01° (binocular viewing, recording right eye only). Each sentence was presented on a single line in 14-point Courier New font on a 19-in LCD monitor positioned 55 cm in front of the participants (1° visual angle \approx 3 characters). The eye tracker was calibrated before beginning the experiment and subsequently was recalibrated between trials as necessary.

Regions of Analysis and Data Processing. Each trial was inspected by hand using the University of Massachusetts EyeDoctor software suite (37). We discarded any trial in which there was track loss before some fixation in any region other than the final region. This process resulted in loss of 15.3% of trials. Most of these track losses were caused by the participant blinking.

We examined a number of standard eye movement measures (33) including: (i) the frequency with which a region was skipped on first reading, (ii) first fixation duration (the duration of the first fixation on a region when no material to the right of the region had yet been fixated), (iii) first pass reading time (the total fixation time on a region the first time it was entered, when no material to the right of the region had yet been fixated; also called gaze duration for regions consisting of only one word), (iv) go-past time (the accumulated time from when a reader first fixated on a region until their first fixation to the right of the region; this measure includes any regressions the reader made before moving forward past the word), (v) total reading time (the summed time of all fixations on a region), (vi) regressions out of a region immediately after first-pass reading, and (vii) regressions into a region. These measures were computed for the regions of each sentence as delineated in example 4. In addition, we defined a go-past regression to have occurred from region Y to region X if the reader had a first-pass regression from region Y and subsequently fixated on region X before saccading past region Y. Fig. 2C shows frequencies of go-past regression from the critical region to the preposition.

Statistical Analysis Method. We report most results using traditional by-participants (F1) and by-items (F2) ANOVAs, with proportions for binary-valued outcomes arcsine transformed before statistical analysis. Consistent with standard practice, reading times >4 SDs outside the mean for each condition in each region were discarded as outliers. In cases where the assumptions of ANOVA were badly violated (heavily imbalanced data and/or binary responses with by-subject or by-item means close to 0 or 1), we used mixed-effects models with crossed random effects of subject and item (38) using the lme4 package in R (39). Complete details can be found in *SI Appendix*.

ACKNOWLEDGMENTS. We thank E. Bakovic for assistance with norming studies, L. Frazier for helpful discussion, and E. H. Levy and E. Mukamel for helpful feedback on the manuscript. This work was supported by National Institutes of Health Grant HD26765 and the University of California at San Diego Academic Senate.

- McClelland JL, Elman JL (1986) The TRACE model of speech perception. *Cognit Psychol* 18:1–86.
- Marslen-Wilson WD (1987) Functional parallelism in spoken word recognition. *Cognition* 25:71–102.
- Grosjean F (1985) The recognition of words after their acoustic offset: Evidence and implications. *Percept Psychophys* 38:299–310.
- Bard EG, Shillcock RC, Altmann GTM (1988) The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Percept Psychophys* 44:395–408.
- Warren RM, Warren RP (1970) Auditory illusions and confusions. *Sci Am* 223:30–36.
- Connine CM, Blasko DG, Hall M (1991) Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *J Mem Lang* 30:234–250.
- Dahan D, Tanenhaus MK (2004) Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *J Exp Psychol Learn Mem Cognit* 30:498–513.
- McMurray B, Tanenhaus MK, Aslin RN (2009) Within-category VOT affects recovery from “lexical” garden-paths: Evidence against phoneme-level inhibition. *J Mem Lang* 60:65–91.
- Fodor JA (1983) *The Modularity of Mind* (MIT Press, Cambridge, MA).
- Jurafsky D (1996) A probabilistic model of lexical and syntactic access and disambiguation. *Cognit Sci* 20:137–194.
- Hale J (2001) A probabilistic Earley parser as a psycholinguistic model. *Proc NAACL* 2:159–166.
- Levy R (2008) Expectation-based syntactic comprehension. *Cognition* 106:1126–1177.
- Reichle ED, Pollatsek A, Fisher DL, Rayner K (1998) Toward a model of eye movement control in reading. *Psychol Rev* 105:125–157.
- Engbert R, Nuthmann A, Richter EM, Kliegl R (2005) SWIFT: A dynamical model of saccade generation during reading. *Psychol Rev* 112:777–813.
- Levy R (2008) A noisy-channel model of rational human sentence comprehension under uncertain input. *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pp 234–243.
- Slattery T (2009) Word misperception, the neighbor frequency effect, and the role of sentence context: An eye movement investigation. *J Exp Psychol Hum*, in press.
- Tabor W, Galantucci B, Richardson D (2004) Effects of merely local syntactic coherence on sentence processing. *J Mem Lang* 50:355–370.

18. Gibson E (2006) The interaction of top-down and bottom-up statistics in the resolution of syntactic category ambiguity. *J Mem Lang* 54:363–388.
19. Van Dyke JA (2007) Interference effects from grammatically unavailable constituents during sentence processing. *J Exp Psychol Learn* 33:407–430.
20. Bever T (1970) The cognitive basis for linguistic structures. *Cognition and the Development of Language*, ed Hayes J (Wiley, New York), pp 279–362.
21. Konieczny L (2005) The psychological reality of local coherences in sentence processing. *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, eds Bara BG, Barsalou L, Bucciarelli M, (Lawrence Erlbaum Associates, Inc. Mahwah, NJ), pp 1178–1183.
22. Frazier L, Fodor JD (1978) The sausage machine: A new two-stage parsing model. *Cognition* 6:291–325.
23. Tanenhaus MK, Spivey-Knowlton MJ, Eberhard K, Sedivy JC (1995) Integration of visual and linguistic information in spoken language comprehension. *Science* 268:1632–1634.
24. Altmann GT, Kamide Y (1999) Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73:247–264.
25. Staub A, Clifton C (2006) Syntactic prediction in language comprehension: Evidence from *either ... or*. *J Exp Psychol Learn* 32:425–436.
26. MacDonald MC (1994) Probabilistic constraints and syntactic ambiguity resolution. *Lang Cognit Processes* 9:157–201.
27. Spivey MJ, Tanenhaus MK (1998) Syntactic ambiguity resolution in discourse: Modeling the effects of referential content and lexical frequency. *J Exp Psychol Learn* 24:1521–1543.
28. McRae K, Spivey-Knowlton MJ, Tanenhaus MK (1998) Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *J Mem Lang* 38:283–312.
29. Narayanan S, Jurafsky D (2002) A Bayesian model predicts human parse preference and reading time in sentence processing. *Adv Neural Information Processing Syst* 14:59–65.
30. Cover T, Thomas J (1991) *Elements of Information Theory* (Wiley, New York).
31. Kucera H, Francis WN (1967) *Computational Analysis of Present-Day American English* (Brown Univ Press, Providence, RI).
32. Marcus MP, Santorini B, Marcinkiewicz MA (1994) Building a large annotated corpus of English: The Penn Treebank. *Comput Linguist* 19:313–330.
33. Rayner K (1998) Eye movements in reading and information processing: 20 years of research. *Psychol Bull* 124:372–422.
34. Legge GE, Klitz TS, Tjan BS (1997) Mr. Chips: An ideal-observer model of reading. *Psychol Rev* 104:524–553.
35. Najemnik J, Geisler WS (2005) Optimal eye movement strategies in visual search. *Nature* 434:387–391.
36. Griffiths TL, Tenenbaum JB (2006) Optimal predictions in everyday cognition. *Psychol Sci* 17:767–773.
37. Stracuzzi D, Kinsey J (2009) *Eyetrack Software for the SR Eyelink II and Eyelink 1000, Version 0.7.10k* (University of Massachusetts Amherst, Amherst, MA).
38. Baayen RH, Davidson DJ, Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang* 59:390–412.
39. Bates D (2005) Fitting linear mixed models in R. *R News* 5:27–30.