Journal of Experimental Psychology:
Learning, Memory, and Cognition
2010, Vol. 36, No. 2, 538–544

0278-7393/10/$12.00 DOI: 10.1037/a0018027

RESEARCH REPORTS

Effects of Plausibility on Structural Priming

Kiel Christianson and Steven G. Luke
University of Illinois at Urbana–Champaign

Fernanda Ferreira
University of Edinburgh

We report a replication and extension of Ferreira (2003), in which it was observed that native adult English speakers misinterpret passive sentences that relate implausible but not impossible semantic relationships (e.g., *The angler was caught by the fish*) significantly more often than they do plausible passives or plausible or implausible active sentences. In the experiment reported here, participants listened to the same plausible and implausible passive and active sentences as in Ferreira (2003), answered comprehension questions, and then orally described line drawings of simple transitive actions. The descriptions were analyzed as a measure of structural priming (Bock, 1986). Question accuracy data replicated Ferreira (2003). Production data yielded an interaction: Passive descriptions were produced more often after plausible passives and implausible actives. We interpret these results as indicative of a language processor that proceeds along differentiated morphosyntactic and semantic routes. The processor may end up adjudicating between conflicting outputs from these routes by settling on a "good enough" representation that is not completely faithful to the input.

Keywords: good-enough language processing, structural priming, plausibility, syntactic parsing, language comprehension

Imagine hearing a sentence like *The dog was bitten by the man*. There are at least two ways to arrive at an interpretation of this sentence. One is to build a syntactic structure strictly according to the morphosyntactic rules of English applied to the lexical items in the sentence, irrespective of any semantic knowledge one might have about those lexical items. Another is to examine the semantics of the individual lexical items and map them onto semantic roles such as agent and patient according to one's experience with those lexical items; that is, dogs usually bite people rather than the other way around. Whereas the former option will invariably lead to the correct interpretation, it may not be the appropriate option in all circumstances, for example, when the listener has reason to believe that the speaker has misspoken or when the listener is a nonnative English speaker with limited English morphosyntactic knowledge.

An interesting question is to what extent each of these two routes toward comprehension is pursued by adult native speakers during normal language comprehension. If processing proceeds along just one route regardless of the communicative situation, the problem arises of how native and nonnative speakers comprehend language under less-than-optimal conditions. If processing naturally proceeds along more than one route, however, it is not clear whether integration of the outputs from the two routes is mandatory, or, if it is not, how the language comprehension system ultimately decides which of the outputs is the right one to use. Note that for the example given here, the first option returns a different interpretation from the second, whereas if the sentence instead were *The man was bitten by the dog*, both routes would lead to the same interpretation.

Recent electrophysiological studies of language comprehension have provided evidence of the existence of two distinct processing routes, as sketched above (Kim & Osterhout, 2005; Kuperberg, 2007; Kuperberg, Caplan, Sitnikova, Eddy, & Holcomb, 2006; Schlesewsky & Bornkessel, 2006; van Herten, Chwilla, & Kolk, 2006). Kim and Osterhout (2005) collected event-related potential (ERP) data from people reading sentences such as *The hearty meal was devouring the kids* and observed an inflated positive deflection at approximately 600 ms (i.e., a P600 response) after the verb, compared with active and passive controls. This P600 response is typically associated with syntactic reanalysis but was observed despite the unambiguous syntactic information in the sentences. Kim and Osterhout observed no significant negative deflection at 400 ms (N400), which is typically interpreted as evidence of

Kiel Christianson, Department of Educational Psychology and Beckman Institute for Advanced Science and Knowledge, University of Illinois at Urbana–Champaign; Steven G. Luke, Department of Educational Psychology, University of Illinois at Urbana–Champaign; Fernanda Ferreira, Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom.

This research was funded in part by a grant to Kiel Christianson from the Campus Research Board at the University of Illinois at Urbana–Champaign and generous support from the Department of Educational Psychology, University of Illinois at Urbana–Champaign.

We thank Kent Lee, Jeong-Ah Shin, Jung Hyun Lim, Hee Youn Cho, Ji Kim, Michael Blasingame, Eric Juul, Allie Stanko, Rachel Leddy, Emily Hayes, Cindy Johnson, Jennifer Brown, Amelia Coleman, and the rest of the members of the Educational Psychology Psycholinguistics Lab at the Beckman Institute for assistance in running participants and coding production data. Thanks also to Ted Gibson for invaluable comments on earlier versions of the article.

Correspondence concerning this article should be addressed to Kiel Christianson, Department of Educational Psychology, Education Building, Room 226A, MC-708, University of Illinois, 1310 South 6th Street, Champaign, IL 61820. E-mail: kiel@illinois.edu

semantic anomaly detection. Kim and Osterhout interpreted these data as demonstrating that in some cases, "semantic processing operates independently of and perhaps even controls syntactic analysis" (Kim & Osterhout, 2005, p. 210).

Consistent with Kim and Osterhout's (2005) conclusion, Kuperberg (2007) also argued that "normal language comprehension proceeds along at least two competing neural processing streams," (p. 23), namely, a semantic memory-based stream and a combinatorial morphosyntactic stream. Kuperberg also observed that implausible sentences are sometimes repaired by swapping the thematic roles of the verb's arguments (cf. Kuperberg et al., 2006).

Van Herten et al. (2006) also presented data consistent with Kim and Osterhout (2005) and with Kuperberg's (2007) dual stream view. Again using ERPs, van Herten and colleagues (see also Kolk, Chwilla, van Herten, & Oor, 2003) examined participants' interpretations of sentences containing implausible strings such as *The fox that shot the poacher.* For such sentences, van Herten et al. found no N400 effect, which would have signaled a semantic anomaly. Instead, they observed P600 effects, signaling an apparent revision of the syntax into a structure that would accommodate a plausible interpretation of that string of words (e.g., a passive structure). They concluded that language processing proceeds along a heuristic route, which takes word order and plausibility into account, and along an algorithmic route, which computes morphosyntactic structure. On the basis of their results, they proposed that the language processor "attempts to combine *all* items into a single representation" (van Herten et al., 2006, p. 1194). Taken together, these electrophysiological results support the existence of parallel independent but interactive syntactic and semantic processing routes. Behavioral evidence for these distinct processing routes or streams has been sparse; however, one study by Ferreira (2003) provided evidence that plausibility information and morphosyntactic information compete during comprehension.

Ferreira (2003) examined the comprehension of *noncanonical* sentences, such as the passive-voice example above, compared with *canonical* English sentences. Noncanonical sentences were passive sentences and object-cleft sentences (*It was the man that the dog bit*); canonical sentences were active (*The dog bit the man*) and subject-cleft sentences (*It was the dog that bit the man*). All sentence types were manipulated such that they communicated a plausible, reversible, but semantically biased transitive event (*dog biting man*); an implausible semantically biased event (*man biting dog*); a nonreversible, anomalous event (*cheese eating mouse*); or a fully reversible event (*woman visiting man*). After hearing each sentence, participants were asked to decide either who the thematic agent (the instructions referred to the "do-er") was or who the thematic patient or theme (the "acted-on") was. (There were four other questions for filler items.) The basic results of the three experiments reported by Ferreira was that people were more accurate in their comprehension for plausible passives than they were for implausible passives (and object clefts, as well), but plausibility did not affect accuracy for actives (or subject clefts).

Ferreira (2003) interpreted these results as evidence for a specific instantiation of a two-route model. The model distinguishes between a heuristic route, on the one hand, and an algorithmic (syntactic) route, on the other. The heuristic route makes use of two heuristics: a word order heuristic, which holds that noun–verb–noun (NVN) strings are mapped initially onto an agent–verb–patient semantic template, and a semantic plausibility heuristic, which states that the processor adopts the semantic analysis most consistent with world knowledge. Simple modeling of the data from her three experiments demonstrated that both heuristics are necessary to account for the entire set of findings, with the word order heuristic weighted more heavily than the semantic heuristic. The algorithmic route was not examined in the experiments, but Ferreira assumed the need to include a syntactic, algorithmic route to account for the overall accuracy rates: That is, although the point of her studies was to show that even simple passive and active sentences are often misinterpreted, most of the time participants did assign thematic roles correctly. Clearly, then, the output of the heuristic route is coordinated with an algorithmic route that usually wins in the competition to assign the sentence a final interpretation. Both of the heuristics Ferreira proposed are essentially semantic: The NVN strategy focuses on assignment of thematic roles, and the other heuristic makes use of semantic plausibility. For the remainder of this article, we do not distinguish between them but treat both heuristics as part of the operation of the semantic route.

In an attempt to provide corroborative evidence of the hypothesized semantic and syntactic processing routes and perhaps the influence of task on both, we combined a slightly modified version of Ferreira's auditory comprehension paradigm with a structural priming paradigm (Bock, 1986; Pickering & Ferreira, 2008). The paradigm thus contained both an offline, explicit comprehension measure and an implicit measure of structural computation. The predicted result for the comprehension question accuracy was a replication of the effects observed by Ferreira (2003). Namely, we expected more misinterpretations of implausible sentences overall, with a potentially greater effect for passives than actives. Predictions for the structural priming results were as follows. If, as van Herten et al. (2006) argued, the language processor tries to integrate the outputs of all processing routes into a single representation, we should observe evidence of the semantic plausibility of the input sentence on the syntactic structure of subsequent production. In other words, when semantic route and syntactic route outputs are integrated, we should observe interactive effects. Specifically, given that semantic representations of sentences appear to be relatively resistant to revision (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Christianson, Williams, Zacks, & Ferreira, 2006; Sturt, 2007), we might expect priming of passive structures only when passive primes are plausible, as implausible passives might be reprocessed to arrive at a structure that is consistent with the semantic output. Furthermore, we might even expect semantic effects on production subsequent to implausible active sentences for the same reason. This latter result is one that has not previously been reported in the literature, yet it would provide considerable evidence for the existence and interaction of the two processing routes described above.

Method

Participants

Seventy-five people recruited from the University of Illinois at Urbana–Champaign Educational Psychology subject pool participated. Participants were compensated for their time with either course credit or $7.

Materials

The 24 semantically biased sentences from Ferreira (2003) were used. These sentences were chosen because they produced on the whole the highest rate of misinterpretation in the previous study, and it was critical for the present experiment to have enough incorrect trials to analyze and compare with correct trials. The sentences, which had been normed for plausibility and reversibility by Ferreira, were recorded digitally by a male native speaker of American Midwestern English at a rate of 10 kHz on the Praat program (Boersma & Weenink, 2005). Two hundred thirty-one filler items were also digitally recorded, which consisted of intransitive active sentences, locative sentences, and existential sentences. Approximately one third of the fillers contained two or more nominal referents. Fillers were generally plausible, but approximately 20% of them were designed specifically to keep participants on their toes. These included implausible items (e.g., *The artist cooked dinner and the chef painted the flower*, *The basketball player was short and the golfer was tall*), temporarily ambiguous structures (with unhelpful prosody) in which implausibility may aid in ambiguity resolution (e.g., *As the old woman mended the sun set*, *As the goalkeeper punted the referee collapsed*), temporarily ambiguous structures in which plausibility did not aid in resolution (e.g., *Sandra tipped the waiter and the hostess screamed*), and factual sentences with somewhat bizarre content (e.g., *Madonna wrote a children's book*, *Wal-Mart sells guns but not rap CDs*). Crucially, all comprehension questions in both fillers and experimental items referred to the thematic roles or locative relationships of the noun phrases in the sentences. Therefore, even in fillers such as *The television was next to the window*, which is plausible but fully reversible, participants would need to pay close attention to the attributes and actions of the noun phrases to answer the questions.

Experimental items were constructed according to a 2 (voice: active vs. passive) × 2 (plausibility: plausible vs. implausible) design. Comprehension questions were constructed in a form designed to avoid priming any passive or active structure and thus interfering with subsequent picture descriptions (as was found to be an issue in extensive pilot testing). These questions contained only the verb from the previous sentence suffixed with either an agentive *–er* or patient *–ee*, an equal (=) sign, and one of the nouns from the sentence, followed by a question mark. Participants answered by pressing either a button marked *YES* or a button marked *NO* on a button box. The questions thus inquired as to the identity of the agent or the patient or theme, as in Ferreira's (2003) original study. A schematic representation of one trial's presentation is shown in Figure 1.

The noun that appeared in the question was counterbalanced across lists, such that in half of the items, it was the actual patient noun phrase, and in the other half, it was the actual agent noun phrase. This yielded a 2 (voice) × 2 (plausibility) × 2 (question type: agent vs. patient) design. Eight lists were constructed such that each item appeared in one and only one version on each list, according to a Latin square design. The order of the experimental items and fillers was pseudorandomized within the following constraint: A minimum of 10 filler items intervened between experimental items to avoid carryover from one experimental item to the next of any structural priming effect (K. Bock, personal communication, January 2006). All fillers appeared on each list.

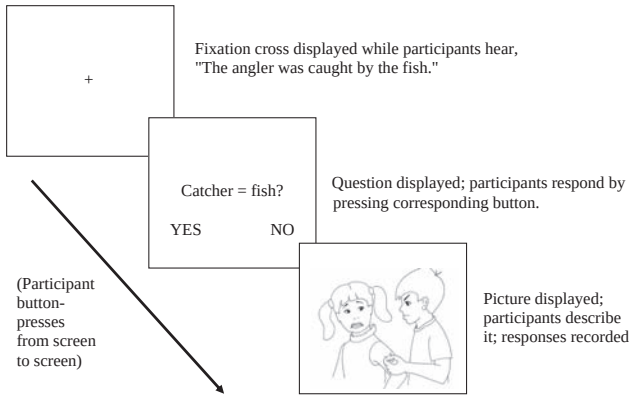

Figure 1. Schematic representation of trial procedure.

An example item and a comprehension question from each condition are given in Table 1.

Line drawings of transitive events were presented after experimental items. All drawings had been used in previous language production experiments (Bock, 1986; Christianson, 2002; Christianson & Ferreira, 2005; Griffin & Bock, 2000; Loebell & Bock, 2003) and were used with permission. Line drawings following filler items depicted various intransitive, locative, and existential scenes and also were taken from the same sources with permission. Several of the drawings included scenes that were implausible or clearly fictitious, for example, a tiny truck on top of a giant book, a large bag of money falling on a man's head, and a baby tickling a porcupine.

Procedure

After granting informed consent, participants were seated in front of a computer monitor. Instructions were given orally and in writing on the computer screen. Participants were then fitted with a lightweight headset consisting of earphones and a microphone. The headset microphone was connected to a Sony ICD-ST25 digital microrecorder, which was switched on to record all vocalizations from the beginning of six practice items (which could be repeated if necessary until participants felt comfortable with the procedure) through the end of the experiment. The experiment was run on E-Prime software, and question responses and reaction times were recorded automatically on an E-Prime button box. Each experimental session was preceded by six practice items, which were repeated as many times as necessary to ensure comprehension of the procedure. Oral production data for experimental items were later transcribed by two native English-speaking research assistants.

Results

Comprehension Question Results

Data from three participants were discarded because the participants produced almost no complete sentences in the production task. We analyzed the data for the remaining 72 participants. The accuracy data are summarized in Table 2. These data were analyzed using a binomial logit mixed model (Jaeger, 2008) in R (R

Table 1
Example of Test Item and Comprehension Question in Each Condition

Condition	Example
Item	
Plausible active	The angler caught the fish.
Implausible active	The fish caught the angler.
Plausible passive	The fish was caught by the angler.
Implausible passive	The angler was caught by the fish.
Question	
Agent	Catcher = angler (fish)?
Patient	Catchee = angler (fish)?

Development Core Team, 2008). The accuracy rate on filler items was 88%, suggesting that participants were attending to the items. Plausibility, voice, and question type were included as fixed effects in the model, with subject and item as random effects (random intercepts only). There was an effect of plausibility, with accuracy rates significantly higher for plausible than for implausible sentences (0.88 vs. 0.79; estimate = −2.85, *SE* = 1.093, $z = -2.8$, $p < .01$). There was also an effect of voice such that active sentences were responded to more accurately than were passives (0.86 vs. 0.80; estimate = −2.96, *SE* = 1.09, $z = -2.9$, $p < .01$). The effect of question type was also significant, indicating that participants were more accurate when answering agent questions versus patient questions (0.88 vs. 0.78; estimate = −4.17, *SE* = 1.075, $z = -4.1$, $p < .001$).

Some interactions also reached significance. Plausibility interacted with question type (agent *–er* vs. patient *–ee*; estimate = 2.83, *SE* = 1.13, $z = 2.65$, $p < .01$). This interaction indicates that the estimate of the effect of plausibility on question accuracy, which was significant for agent questions, must be adjusted for patient questions by adding to it the estimate of the interaction. This reduces the estimate to almost zero (0.02) for patient questions (Baayen, 2008). In other words, the effect of plausibility is not significant for the patient questions. Voice and question type also interacted (estimate = 3.37, *SE* = 1.13, $z = 3.15$, $p < .01$). This indicates that, like the effect of plausibility, the effect of voice that was present for the agent questions disappeared for the patient questions.

These interactions show that the voice and plausibility only affected accuracy for agent questions. To investigate this further, we separated the data by question type and analyzed agent and patient questions separately. For agent questions, the effect of voice was significant (estimate = −3.12, *SE* = 1.15, $z = -2.7$, $p < .01$), indicating that active sentences were responded to more accurately than were passive sentences. There was also an effect of plausibility for agent questions (estimate = −2.98, *SE* = 1.15, $z = -2.6$, $p < .01$), with accuracy rates significantly higher for plausible than for implausible sentences. For patient questions, neither of these effects was significant (all zs < -1.8).

The question accuracy results reported here generally replicate Ferreira (2003). As in the present experiment, Ferreira observed main effects of voice and plausibility, with accuracy greatest in the active and plausible conditions, respectively. The significant interaction of voice and question type reported here was also observed by Ferreira. These findings replicate Ferreira's findings that participants often misinterpret passive sentences and that they do so even more frequently when a passive sentence expresses an implausible idea. Some differences in the findings do exist, however. Ferreira reported a significant interaction of voice and plausibility and a significant three-way interaction, neither of which were significant here (for Voice × Plausibility, $z = 1.36$; for the three-way interaction, $z = -1.77$). The effect of question type and the interaction of plausibility and question type in the present study were not significant in Ferreira's study. These differences are likely the result of methodological differences between the two studies, specifically, differences in the comprehension task. Ferreira's participants were presented with a prompt and generated a free response, whereas in the present experiment, participants were required to make a categorical *yes* or *no* response to an essentially asyntactic question (see Table 1).

The failure to find an effect of voice or plausibility on response accuracy for patient questions may be due to hypothesized tighter links between agents and intrinsic semantic properties (e.g., animacy) and discourse prominence (topic or focus status). Christianson (2002) and Christianson and Cho (2009) observe that expectations about the properties of upcoming or pro-dropped (elided) arguments appear to be more elaborated for agents than for patients. This tentative explanation requires further examination, however.

Production Data

The production data were also analyzed using a binomial logit mixed model. Voice, plausibility, question type, and accuracy were included as fixed effects in the model, with subject and item as random effects (random intercepts only). There were no significant effects of voice, plausibility, question type, or question accuracy (all zs < 1.59). There was a significant interaction of plausibility and voice (estimate = −3.36, *SE* = 1.36, $z = -2.45$, $p < .05$), such that the effect of plausibility was opposite in the two voice conditions: A greater proportion of passive (vs. active) sentences was produced after implausible primes in active voice, whereas a greater proportion of passive (vs. active) sentences was produced after plausible primes in passive voice (see Figure 2). (Note that results in Figure 2 are given in proportions, although logit mixed models are computed over individual data points. We feel proportional data are easier to visualize in graphical form.) All other interactions were nonsignificant (all zs < 1.86). This finding indicates that, at least in some cases, it was not just the actual syntactic structure of the prime sentence that primed subjects' production. Instead, the syntactic structure that was most plausible

Table 2
Mean Comprehension Question Accuracy

	Agent		Patient	
Item condition	*M*	*SD*	*M*	*SD*
Active				
Plausible	0.995	0.07	0.78	0.42
Implausible	0.91	0.28	0.77	0.42
Passive				
Plausible	0.90	0.30	0.83	0.38
Implausible	0.74	0.44	0.75	0.44

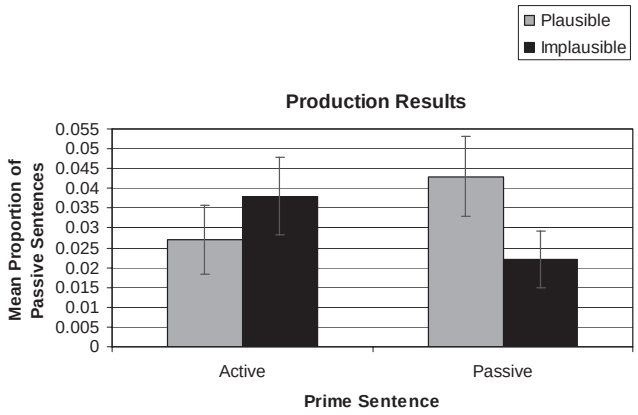

Figure 2. Mean proportion of passives produced by all participants. Error bars represent standard errors.

given the word order seems to have primed subjects' production. For plausible sentences, passive primes led to more passive sentences produced, exactly the kind of priming observed in previous structural priming studies (cf. Pickering & Ferreira, 2008). For implausible sentences, however, active sentences primed more passive sentences, suggesting that at some point the structure was flipped (active to passive and vice versa) to make it more believable. We discuss the implications of this finding in the Discussion section.

Question response accuracy was included as a factor in the linear mixed effects analysis of production data. Neither the main effect of accuracy nor any interaction involving accuracy approached significance. Therefore, the interaction of voice and plausibility reported above seems to be unrelated to accuracy. To explore this null result, we excluded from the production data all responses that had been preceded by an incorrect answer and reran the linear mixed effects analysis. Even with only the correct trials included, the results were the same: The only significant result was the interaction of voice and plausibility (estimate $= -3.41$, $SE = 1.38$, $z = -2.46$, $p < .05$). Additionally, we ran a post hoc linear mixed effects analysis removing the factor of question type and thereby increasing the number of observations per condition. Again, we observed no relationship between the question accuracy and production data (all ps $> .25$). The results of these analyses suggest that the lack of any effect of accuracy on production is not due to low statistical power. Instead, it appears that the type of structure produced is independent of the response given to the comprehension question. We return to this issue in the Discussion section.

Discussion

A growing number of studies suggest that language comprehension proceeds along two routes or streams: a morphosyntactic route and a semantic route. Similarly, Jackendoff's (2007) parallel architecture for language processing, Bornkessel and Schlesewsky's (2006) extended argument dependency model, and Clahsen and Felser's (2006) shallow structure hypothesis all propose parallel routes toward comprehension, with each route delivering its own output. When these outputs converge, language processing and ultimate comprehension are simple. When the outputs conflict—for example, when the morphosyntactic route outputs a representation that is semantically implausible—language comprehension becomes more difficult, and misinterpretations arise. The data presented here are important because they suggest that such misinterpretations are not due to random error or general confusion. Rather, they are due to the language processor's attempts to integrate the outputs from the different processing routes (cf. van Herten et al., 2006).

Consistent with this view, Ferreira (2003) observed that comprehenders seem to make use of a semantic plausibility heuristic and a surface-order-to-thematic-template mapping heuristic as well as syntax to derive meaning from language input. The comprehension question accuracy results reported here basically replicate those of Ferreira, in which noncanonical morphosyntax (i.e., passive structure) led to inflated error rates in the offline questions when that structure conflicted with the plausibility of the sentence. However, it must be stressed that even when both word order and plausibility point away from the correct interpretation (as in the passive, implausible condition), accuracy rates were better than chance (about 74%), indicating that the language processor is generally quite good at resolving the conflicting information sources, as also stressed by Ferreira (2003). Furthermore, the resolution of conflicting semantic and syntactic information, as suggested by van Herten et al. (2006), appears to be an attempt to integrate the semantic route and syntactic route outputs rather than setting one aside in favor of the other. If the morphosyntactic output had been set aside in favor of the semantic output, then the production data should have revealed no syntactic priming effect. If the semantic output had been set aside in favor of the morphosyntactic output, then a standard syntactic priming effect should have been observed, such that the presentation of passive sentences should have primed the production of passive sentences, regardless of factors such as plausibility or accuracy.

Instead, there was an interaction between voice and plausibility in the production data such that passive primes led to more passive sentences in the plausible condition, but, in the implausible condition, active sentences primed passive structures. We interpret this result as indicative of a concerted effort on the part of the language processor to reconcile the outputs of the syntactic and semantic processing routes. When plausibility and morphosyntax converged, the actual syntax of the input sentence was primed for subsequent production. When the input sentence was implausible, though, in at least some of the trials, the semantics of the individual words primed a structure that agreed with the semantic route output. This explains why in the implausible condition there were more passive sentences after active primes and fewer passive sentences after passive primes. This result is completely consistent with the ERP studies cited above.

The question arises as to the mechanism whereby implausible active input would structurally prime passive sentences in production and, conversely, how implausible passive input would fail to prime passive sentences in production. One possibility is that the language processor attempted full syntactic reanalysis of the input sentence when a conflict between the semantic and morphosyntactic route outputs was detected. If this were to occur, it would lead to the prediction that the production data and the question accuracy data should be closely correlated, which they were not. (Below, we discuss this lack of correlation further.) Another pos-

sibility is that in implausible conditions, the language processor performs a less complete thematic reassignment for the arguments. Kuperberg and colleagues (Kuperberg et al., 2006; Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007) presented ERP data suggesting that conflicts between semantic and syntactic route outputs can be addressed via a reassignment of thematic roles to the arguments, possibly without structural reanalysis. Chang, Bock, and Goldberg (2003) reported structural priming data showing that when syntactic structure is held constant, participants tend to produce sentences in which the order of the thematic roles is the same as that of the prime sentence. Thus, if thematic roles were reassigned such that the patient came before the agent in the prime—even without the recomputation of a passive structure—we would expect the patient to precede the agent in the produced sentences as well. This is what Chang et al. observed and is also consistent with the production data reported here.

It should also be noted that results from other experimental paradigms provide converging evidence of final representations of linguistic input that are less than complete, unfaithful to the input, or even syntactically unlicensed. Schlesewsky and Bornkessel (2006) reported ERP data suggesting that in some German dative constructions, reanalysis is achieved under certain circumstances not by structural reconfiguration but rather by "argument reindexation" (p. 139), resulting in structures that are licit elsewhere in the language but not in the constructions used in that study. Christianson et al. (2001, 2006) presented offline question accuracy data suggesting that when faced with certain temporarily ambiguous (garden-path) sentences such as *While Anna dressed the baby spit up on the bed*, people appear to simultaneously maintain both the partial interpretation (that Anna dressed the baby) and the ultimately correct global interpretation (that Anna dressed herself, and the baby spit up on the bed). Tabor, Galantucci, and Richardson (2004) similarly reported that participants in several experiments experienced significant disruptive effects of "merely local syntactic coherence" such as "the player tossed the Frisbee" in the sentence *The coach smiled at the player tossed the Frisbee*. Crucially, the apparent coherence of the strings in question should have been ruled out by the preceding syntax.

All of the above results, along with those reported here, can be attributed to language processing that has been described as "good enough" by Ferreira and colleagues (Ferreira, Bailey, & Ferraro, 2002; Ferreira, Christianson, & Hollingworth, 2001; Ferreira & Patson, 2007). Good-enough language processing results in a final representation that is underspecified in some way, such that it is unfaithful to the input. This underspecification is hypothesized to result from the failure to successfully integrate the outputs of various processing routes in cases where the outputs conflict. Semantic plausibility appears to be a driving force in the success of this final integration. Christianson et al. (2001) showed that misinterpretations were significantly reduced for garden-path sentences such as *While Anna dressed the baby splashed in the tub*, the intermediate, partial interpretation (that Anna dressed the baby) of which is rendered implausible by subsequent material. Furthermore, it has been proposed that good-enough processing, as a least-effort process, is affected by task demands, cognitive load, and individual differences in cognitive function (Ferreira & Patson, 2007; Swets, Desmet, Clifton, & Ferreira, 2008). As such, integration can reasonably be expected to result in a final representation that fails to fully represent all of the information contained in the input (as also is the case in other areas of cognition, e.g., visual processing; cf. Ferreira et al., 2001). The results presented here further explicate good-enough processing. Specifically, the conflict between semantic and morphosyntactic route outputs can be adjudicated by the processor in such a way as to create syntactic structures (or at least thematic relations) that reflect semantically plausible relations between arguments and their verbs rather than the actual morphosyntax of the input. In this way, these results are quite similar to those of Kuperberg et al. (2006, 2007), Kim and Osterhout (2005), Schlesewsky and Bornkessel (2006), and van Herten et al. (2006).

Of course, it is not clear yet if the mere existence of a conflict between the semantic and morphosyntactic outputs is enough to trigger an attempt to fully integrate the outputs. Perhaps the language comprehension system is comfortable with the inconsistency until a question is asked, so that asking a question triggers the integration (Christianson et al., 2006). Furthermore, the time course of the integration is not clear, nor is whether the integrative process causes the initial outputs to be automatically overwritten or whether they might continue to exert influence on memory for the input (Sturt, 2007).

Two curious results of the present experiment remain to be addressed. The first is that the priming effects obtained here are relatively small, compared with those generally reported in the structural priming literature. For now, we can only speculate that the introduction of plausibility as a factor may be to blame. To our knowledge, no previous structural priming experiments have included implausible stimuli. We have proposed that the implausibility of some of the items introduced uncertainty about the actual syntactic structures of those items. This could have led participants to doubt the reliability of morphosyntactic cues across items throughout the experiment, including even the plausible items. If the morphosyntax of items was generally less stable or less unambiguously encoded into the implicit memory of participants, it would be expected that it should have less of an effect on subsequent production.

The second curiosity is that there was no apparent relationship between question accuracy and syntactic priming. It is difficult to infer too much from a null result; however, some speculations can be offered. It is possible that a lack of statistical power is to blame, despite attempts to rule this out (see the Results section). An alternative explanation is that despite the language processor's attempt to integrate conflicting outputs from different processing routes, the initial conflicting outputs continue to exert influence on interpretation. Such influence might be asymmetric: Semantic information might more strongly affect explicit, metalinguistic tasks (e.g., answering comprehension questions), and morphosyntactic information might more strongly affect implicit processes (e.g., those involved in structural priming; cf. Shin & Christianson, 2009). This explanation would appear to be consistent with the task effects on ERP results reported by Kuperberg (2007). A third possibility is that the forms of the questions used here and by Ferreira (2003) were such that they triggered "semantic-thematic attraction" (Kuperberg, 2007, p. 23; see also Kim & Osterhout, 2005). Our questions did use lexical pairs, for example, *catcher = angler?* As such, it is possible that the questions themselves added yet another information source to the trial, causing confusion between the outputs and integration of the input sentences and the content of the questions. Ferreira's comprehension checks, that is,

decisions as to who was the do-er or acted-on, are arguably less likely to cause this sort of interference; however, because we replicated the findings of that study here, it would seem clear that those questions either did or did not interfere to a similar extent with the ultimate representations built for these sentences.

To summarize, the present study replicated Ferreira's (2003) observations of misinterpretation effects for simple passive sentences in English and further clarified the interpretation of those results. More important, it demonstrated effects of plausibility on structural priming. This is the first behavioral evidence that nonsyntactic information (plausibility) affects not only the interpretation of sentences but also the ultimate syntactic analysis thereof. We take these results as indicative of the dynamic, highly interactive nature of morphosyntactic and semantic processing routes during sentence comprehension.

References

Baayen, H. R. (2008). *Analyzing linguistic data.* Cambridge, United Kingdom: Cambridge University Press.

Bock, K. (1986). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 575–586.

Boersma, P., & Weenink, D. (2005). Praat: Doing phonetics by computer (Version 4.3.14) [Computer program]. Retrieved from http://www.praat.org/

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review, 113,* 787–821.

Chang, F., Bock, K., & Goldberg, A. E. (2003). Can thematic roles leave traces of their places? *Cognition, 90,* 29–49.

Christianson, K. (2002). *Sentence processing in a 'nonconfigurational' language.* Unpublished doctoral dissertation, Michigan State University, East Lansing, MI.

Christianson, K., & Cho, H.-Y. (2009). Interpreting null pronouns (*pro*) in isolated sentences. *Lingua, 119,* 989–1008.

Christianson, K., & Ferreira, F. (2005). Conceptual accessibility and sentence production in a free word-order language (Odawa). *Cognition, 98,* 105–135.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42,* 368–407.

Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Misinterpretations of garden-path sentences by older and younger adults. *Discourse Processes, 42,* 205–238.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27,* 3–42.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47,* 164–203.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11,* 11–15.

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research, 30*(1), 3–20.

Ferreira, F., & Patson, N. D. (2007). The "good enough" approach to language comprehension. *Language and Linguistics Compass, 1,* 71–83.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11,* 274–279.

Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain Research, 1146,* 2–22.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52,* 205–225.

Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language, 85,* 1–36.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research, 1146,* 23–49.

Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. (2006). Neural correlates of processing syntactic, semantic and thematic relationships in sentences. *Language and Cognitive Processes, 21,* 489–530.

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language, 100,* 223–238.

Loebell, H., & Bock, J. K. (2003). Structural priming across languages. *Linguistics, 41,* 791–824.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin, 134,* 427–459.

R Development Core Team. (2008). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Schlesewsky, M., & Bornkessel, I. (2006). Context-sensitive neural responses to conflict resolution: Electrophysiological evidence from subject-object ambiguities in language comprehension. *Brain Research, 1098,* 139–152.

Shin, J.-A., & Christianson, K. (2009). Syntactic processing in Korean-English bilinguals: Evidence from cross-linguistic structural priming. *Cognition, 112,* 175–180.

Sturt, P. (2007). Semantic re-interpretation and garden-path recovery. *Cognition, 105,* 477–488.

Swets, B., Desmet, T., Clifton, C., Jr., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition, 36,* 201–216.

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language, 50,* 355–370.

van Herten, M., Chwilla, D. J., & Kolk, H. H. J. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience, 18,* 1181–1197.

Received September 11, 2008
Revision received September 25, 2009
Accepted September 30, 2009 ■