



Published in final edited form as:

J Mem Lang. 2010 November 1; 63(4): 489–505. doi:10.1016/j.jml.2010.08.004.

Effects of event knowledge in processing verbal arguments

Klinton Bicknell^a, Jeffrey L. Elman^a, Mary Hare^b, Ken McRae^c, and Marta Kutas^a

^aUniversity of California at San Diego, La Jolla, CA

^bBowling Green State University, Bowling Green, OH

^cUniversity of Western Ontario, London, Canada

Abstract

This research tests whether comprehenders use their knowledge of typical events in real time to process verbal arguments. In self-paced reading and event-related brain potential (ERP) experiments, we used materials in which the likelihood of a specific patient noun (*brakes* or *spelling*) depended on the combination of an agent and verb (*mechanic checked* vs. *journalist checked*). Reading times were shorter at the word directly following the patient for the congruent than the incongruent items. Differential N400s were found earlier, immediately at the patient. Norming studies ruled out any account of these results based on direct relations between the agent and patient. Thus, comprehenders dynamically combine information about real-world events based on intrasentential agents and verbs, and this combination then rapidly influences online sentence interpretation.

Keywords

sentence processing; psycholinguistics; language comprehension; event knowledge; self-paced reading; event-related potentials

A number of recent studies have shown that comprehenders are sensitive to *thematic fit*, or the plausibility of a noun as an argument of a particular verb (e.g., Kamide, Altmann, & Haywood, 2003; McRae, Spivey-Knowlton, & Tanenhaus, 1998). But what is the mechanism that underlies this effect? On one well-established approach, the verb's lexical representation encodes information about typical fillers of its thematic roles, as well as information about the selectional restrictions that the verb imposes on its arguments. Thus when the verb is accessed it makes available information about appropriate role fillers, so that the interpretation of meaning relies crucially on this information from the verb's lexical representation. Another possibility, however, is that comprehenders dynamically compute an interpretation based on their knowledge of events and situations, relying on all available cues. In what follows we test this hypothesis, focusing on the issue of thematic fit. We begin by reviewing the work on that topic, and show that the full pattern of results is difficult to explain on a retrieval-based account. Instead, we argue that comprehenders rely on their knowledge of typical events and situations as they integrate information provided by not only the verb but other participants mentioned in discourse, in an attempt to generate expectancies¹ about upcoming words.

© 2010 Elsevier Inc. All rights reserved.

Corresponding author: Klinton Bicknell, kbicknell@ling.ucsd.edu, University of California, San Diego, Department of Linguistics #0108, 9500, Gilman Dr., La Jolla, CA 92093-0108.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Thematic Fit

In determining thematic fit, comprehenders clearly exploit semantic information about the nature of a verb's arguments and the specific roles they play. Across verbs, animacy (a selectional restriction that has often been invoked and manipulated) has been shown to influence whether an argument is interpreted as agent or patient (Trueswell, Tanenhaus, & Garnsey, 1994), for example, but more verb-specific information is used as well. Listeners have been shown to make anticipatory looks to a picture of an edible object on hearing the verb *eat* (Altmann & Kamide, 1999), and readers of structurally ambiguous sentences are more likely to interpret *cop* as the agent of the verb *arrest* and *crook* as its patient (McRae et al., 1998). Although the choice of patient in the Altmann and Kamide study might be attributable to lexical selectional information (for example, that *eat* requires an edible patient), it would be more difficult to extend that account to the choice of patient for *arrest*. McRae et al. argued instead that a good patient for an event is one which prototypically plays that role, suggesting that thematic fit more broadly reflects the comprehender's knowledge of events and situations. Indeed, Altmann and colleagues point out that their findings are consistent with the use of general experiential knowledge as well (Kamide et al., 2003).

Kamide et al. (2003) also demonstrated that the thematic fit of an object depended on the other available role-fillers. Thus, listeners viewing a display with pictures of two drinkable objects made anticipatory eye movements to a glass of wine on hearing *The woman will drink*, but to milk while hearing *The baby will drink*. These results suggest that comprehenders anticipate the upcoming patient noun based on a combination of linguistic input about the verb and the agent noun, real-world knowledge about common events, and, in this case, visual input about potential role-fillers. It thus appears that a wide variety of distinct information sources - of different levels of representation and in different modalities - rapidly combine to allow comprehenders not only to interpret the current input, but also to anticipate what is yet to come.

We interpret this pattern of results as indicating that comprehension involves the computation of a representation of the relevant event, and the roles of the participants in that event, based on multiple cues made available in the discourse. Although this is not the account that the Kamide et al. (2003) study was designed to test, we view their results as impressive evidence in its favor. Rayner, Warren, Juhasz, and Liversedge (2004), however, have suggested that the rapid plausibility effect found in Kamide et al. is more a function of the experimental method than the comprehension process itself, and they argue that anticipatory effects were found because the visual world paradigm artificially limits the set of potential role fillers, making the arguments more predictable.

We do not share this view because we are of the opinion that context - whether visual, linguistic, or based on world knowledge - has the function of narrowing the assumed domain of discourse in exactly this way. There are a number of reasons for believing that such context-based predictive effects are likely. In an ERP experiment, for example, Federmeier and Kutas (1999) used context to manipulate the extent to which a specific exemplar of a semantic category was expected as a sentence-final noun. The sentences were then completed in one of three ways: with the expected exemplar, with an unexpected exemplar from the same semantic category, or with an unexpected item from another category. Both types of unexpected nouns elicited larger N400s than did the expected one, consistent with the well-known finding that N400 amplitudes are large for semantic incongruities and inversely related to word offline

¹The experiments presented in this paper do not attempt to distinguish between expectation-based and integration-based accounts of processing difficulty. While the literature on this issue is mixed, we here adopt an expectation-based account, following work by DeLong, Urbach, and Kutas (2005); Van Berkum, Brown, Zwitserlood, Kooijman, and Hagoort (2005); and Wicha, Bates, Moreno, & Kutas (2003).

cloze probabilities (Kutas & Hillyard, 1980; Kutas, Lindamood, & Hillyard, 1984). However, a contextually unexpected final noun from the same category as the expected exemplar elicited an N400 that was intermediate between that in the other two conditions. Federmeier and Kutas argue that this is a predictive effect: as comprehenders read, they generate expectations for an ending consistent with the context, activating features of that concept in semantic memory. To the extent that the item shares those features – more so for the category-related than category-unrelated one – it would be partially active and therefore less unexpected (see also Federmeier, McLennan, De Ochoa, & Kutas, 2002).

In the studies we report here, we test the hypothesis that expectations regarding likely fillers of a patient role rely on event knowledge, and that these expectancies are driven by multiple cues that are dynamically integrated during processing. We use both self-paced reading and ERP measures to test this hypothesis, manipulating the thematic fit of a patient noun by holding constant the verb and changing the agent noun to alter the interpretation of the event being described.

The logic underlying this manipulation is that such interactions among the filler of one thematic role, the verb, and expectancies regarding another thematic role are predicted in the case that they result from the comprehender's active use of event knowledge to construct a situation model. These interactions are less likely to result from a strictly lexical retrieval account at least in part because of the combinatoric challenge they pose: the specifications for the fillers of one role would depend on how another role is filled, resulting in a potentially huge number of contingencies that would need to be stored. Embedding detailed world knowledge in the lexicon, however, is inconsistent with prevailing views of lexical representation (e.g., Jackendoff, 2002). Thus, our goal is to demonstrate that comprehenders' knowledge of events in the world plays a major role in on-line sentence processing. Our more general claim is that we do not merely retrieve meaning from a lexicon, but rather compute it from all available cues, of which event knowledge is one.

Experiment 1

We used self-paced reading to investigate whether the choice of agent noun alters the event that a verb describes, leading comprehenders to anticipate a patient that is plausible for that specific event, rather than for the verb in general. Stimuli were designed to reflect comprehenders' judgments of event-based thematic fit, on the one hand, and to minimize other possible sources of facilitation, on the other.

Method

Participants—Forty-eight University of California, San Diego undergraduates (31 women, 17 men, mean age=21.4, SD=2.6 years) participated in the self-paced reading experiment for course credit. In addition, 47 UC San Diego undergraduates participated in a production norming task, 22 in a free association task, and 44 in a priming task; these three tasks were used to develop the materials for the self-paced reading experiment. All participants received course credit and were native English speakers with normal or corrected-to-normal visual acuity.

Materials—Based on intuition, we selected 50 verbs, each paired with two agent nouns that in our judgment distinctly altered the scenario elicited by the agent-verb combination. The verb *check*, for example, was paired with *journalist* and *mechanic*. Good patients for each agent-verb pair were then elicited using production norms designed to assess comprehenders' knowledge of plausible patients for different events.

Production norms—We created 50 pairs of questions, one pair for each verb, to obtain typical patient nouns given the agent-verb combinations. These were presented to participants in a web-based questionnaire. The questions were presented in two lists such that each participant saw all 50 verbs, but only one question per verb. Thus, one participant would respond to “Who or what does a mechanic typically check?” and another to “Who or what does a journalist typically check?” Each participant saw a different random ordering of the stimuli. Space was available for up to 5 responses per question.

Responses were summed across participants, weighted by their position in the five response slots. That is, the participant’s first response was weighted by 5, the second response by 4, and so on. This procedure produced a ranked list of typical patient nouns for each agent-verb pair. A candidate set of experimental items were chosen based on these rankings. First, however, we tested for normative association when the agents were used as stimuli.

Association norms—The goal of Experiments 1 and 2 was to test whether event knowledge elicited by the combination of agent and verb can facilitate interpretation of a patient. Critically, it was important to rule out the possibility that the agent directly facilitated the patient, without any influence of the verb.

In some theories, facilitation of a lexical item is directly related to response strength in the free association task (McNamara 1992, 1994). Therefore, we collected single-response free associations to our agent nouns. The task was conducted using a web-based questionnaire in which participants were given each agent noun and asked to ‘give the first word that comes to mind that is meaningfully related or strongly associated’ to that word. Any patient noun that was produced as a response to its corresponding agent, even if only one of the 22 participants produced it, was removed from consideration as an experimental item. Of the remaining patients for each agent-verb combination, one was chosen that had a high score in the production norms. In some cases where this was difficult, we used a patient that was semantically similar to one of the high-scoring patients, as long as it was not produced in the association norms. We also removed six verbs for which we felt that the agents did not markedly change the types of patient responses. Finally, we removed 12 verbs for which it was difficult to create a natural-sounding set of sentences. This left 32 verbs, each of which had two agents and two patients connected with it, and for which no agent-patient associations were elicited.

Agent-patient priming—However, numerous studies have demonstrated that priming can be obtained even when prime-target pairs are not associated according to free association norms. For example, Hare, Jones, Thomson, Kelly, and McRae (2009) found priming between typical role fillers for a given event, such as an instrument and the type of things on which it is used (*scissors-hair*), and argued that the facilitation was driven by underlying event-based knowledge rather than an undifferentiated associative relationship between the two words. The goals of the current study are somewhat different, however, and consequently we want to rule out, as far as possible, any independent facilitation of the patient by the agent noun. We therefore conducted a priming study to determine whether our agent nouns facilitate their respective patients. Since our aim was to demonstrate a null effect for these items, we also included a set of materials that have shown facilitation in the past (McRae & Boisvert, 1998) to ensure that priming could indeed be obtained with the participants and procedure that we used.

The 64 congruent agent-patient pairs (*journalist-spelling*, *mechanic-brakes*) and the 32 semantically similar McRae and Boisvert (1998) pairs (*eagle-hawk*) were used in a short SOA (250 ms) priming task. Participants performed lexical decisions on the targets. The experimental details are presented in Appendix B.

Because of recent concerns in the literature about the appropriateness of analyses of variance in psycholinguistics (Baayen, 2004; Baayen, Davidson, & Bates, 2008; Dixon, 2008; Jaeger, 2008), we performed generalized linear mixed-effect regression (Breslow & Clayton, 1993) as well as ANOVAs. The results of the mixed-effect regressions are given in brackets following the results of the ANOVAs. (See the Design section in Appendix B for details.) McRae and Boisvert's items showed a significant 18 ms of facilitation [ANOVA: $F_1(1,42) = 8.67, p < .006$; $F_2(1,30) = 4.31, p < .05$; Mixed effect regression: $t = 2.05, p < .05$]. The mean lexical decision latencies for our 32 experimental agent-patient pairs were 509 ms for both congruent (*journalist-spelling*) and incongruent (*mechanic-spelling*) pairs [ANOVA: both $F_s < 1$; Mixed effect regression $t < 1$]. Thus, there is no evidence that the agents independently facilitate their corresponding patients.

Experimental sentences—The three norming studies resulted in 32 verbs, each occurring in two agent-verb-patient triples, such as *journalist checked spelling* and *mechanic checked brakes*. To create a counterbalanced design, we created two target sentences for each verb-patient combination, such that each patient noun was congruent with the event evoked in one item and incongruent with the event in the other. Thus, there were 64 pairs of target sentences in total, in which one of each pair was congruent and the other was incongruent. The items for *checked* are presented below.

1. checked spelling

Congruent: The *journalist* checked the *spelling* of his latest report to the editor. His editor was a stickler for correct spelling and made everyone else at the paper obsess about it too.

Incongruent: The *mechanic* checked the *spelling* of his latest report about the engine. Not having much schooling, he strived to maintain professionalism at his shop.

2. checked brakes

Congruent: The *mechanic* checked the *brakes* on the car he had gotten earlier that day. It turned out that the rear passenger-side brake was the source of the complaint.

Incongruent: The *journalist* checked the *brakes* on the car he had gotten earlier that day. The car had been acting strangely on the way to his interview.

All critical sentences took the form shown in the examples above: (1) the article *The*; (2) the agent noun; (3) the verb; (4) a determiner; (5) the patient noun; (6–8) the post-patient region, which was identical for both the congruent and incongruent version of each item. After this point, the sentences sometimes varied in order to create a plausible continuation. A follow-up sentence was added to each to create a short story, thereby encouraging readers to pay closer attention to meaning. Note that in all cases, the incongruent sentences made sense: in no case were anomalous sentences used.

Items were rotated across four lists so that each participant encountered an equal number of sentences in the congruent and incongruent conditions, but encountered each verb only once. In addition to the 32 experimental sentence pairs on each list, there were 48 filler pairs, arranged so that experimental trials never occurred in sequence. Each trial was followed by a yes-no comprehension question.

Procedure—Participants read sentences presented on a computer screen in a word-by-word non-cumulative self-paced moving window reading paradigm (Just, Carpenter, & Woolley, 1982). A CMU button box measured reading times with millisecond accuracy. The experiment

was controlled by PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993). Each trial began with the entire target sentence displayed on screen, with a dash (-) replacing each letter. As participants read, they pressed a button to reveal the next word, simultaneously causing the previous word to revert to dashes. The button box recorded the time between the presentation of a word and the button press to bring up the next word. After the last word of the target sentence had been read, the follow-up sentence was presented, again displayed in dashes. The comprehension question was displayed after participants had read the follow-up sentence, and participants answered using specified buttons on the button box. Participants were instructed to read as naturally as possible.

Before the main experiment, participants received 20 practice trials to familiarize themselves with the task. They then received a break prior to the start of the main experiment, and additional breaks after every 21 trials. The entire session lasted approximately 30 minutes.

Design—To minimize the effects of word length differences across conditions and of participant reading rates, a regression was performed on the data for each participant, predicting reading time from word length in characters. This process utilized all words from both filler and target trials. The values predicted by the regressions were subtracted from the actual reading times to produce residual reading times.

Analyses of variance were conducted on the residual reading times and comprehension error data for the one-word regions described below. The factor of interest was congruence (congruent vs. incongruent), which was within subjects (F_1) and items (F_2). List was included as a between-participants dummy variable and item rotation group as a between-items dummy variable to stabilize variance that may result from rotating participants and items over lists (Pollatsek & Well, 1995). Effects involving these dummy variables are not reported. The critical regions, illustrated in the example below, were the patient region (*brakes*), the first post-patient region (*on*), and the second post-patient region (*the*).

Example: The mechanic checked the / brakes / on / the / car he ...

As in the priming study, we also performed generalized linear mixed-effect regression (Breslow & Clayton, 1993). For each region's residual reading time data, we fit a linear mixed-effects regression, including congruence and the logarithm of the word's frequency in the British National Corpus (BNC) as fixed effects. Random participant and item intercepts and random participant and item slopes of congruence were also included. (The model also includes a parameter for the correlation between the random intercept and slope variances.) Each region's error data were analyzed with a generalized linear mixed-effect regression with a logit link function using the same fixed and random effects (except for the BNC frequency data). Further details on our mixed-effects regression procedure are available in Appendix B (in the Design section).

Results

Residual reading times more than three standard deviations from the mean for each combination of position and condition were replaced by the cutoff value (< 1% of the data). The mean residual reading times by word and condition are presented in Table 2 and Figure 1. There was no influence of congruence at the patient noun [ANOVA: $F_1(1,44) = 2.23, p > .1$; $F_2(1,60) = 1.55, p > .2$; Mixed effect regression: $t = 1.19, p > .2$]. However, reading times were shorter in the congruent condition at the word directly following the patient [ANOVA: $F_1(1,44) = 11.50, p < .002$; $F_2(1,60) = 9.56, p < .004$; Mixed effect regression: $t = 2.58, p < .004$] as well as at the next word [ANOVA: $F_1(1,44) = 11.20, p < .003$; $F_2(1,60) = 9.73, p < .004$; Mixed effect regression: $t = 2.87, p < .006$].

It is possible, however, that in modifying the interpretations of the event being described, the event participants also modify the sense of the verb. Many of our verbs do indeed show the type of clear sense difference that could arguably result in distinct lexical representations for each sense. Thus, to test whether our results were due to sense-based lexical information we repeated the analyses, this time including only the 17 verbs that do not show such differences (the items included in the subset analysis are marked as such in Appendix A.) Despite including only half as many items, this subset showed the same pattern found overall: Significant facilitation, and of the same magnitude as the with the larger set, in the congruent condition at the word directly following the patient [-4.8 vs. 17.3 ms; $F(1,44) = 7.54, p < .009$; $F(2,1,30) = 6.65, p < .016$] as well as at the next word [-15.8 vs. 5.0 ms; $F(1,44) = 13.90, p < .001$; $F(2,1,30) = 7.29, p < .012$]. Thus there is no evidence that lexically defined sense differences drive the overall effect. Indeed, the similarity between this and the overall analysis suggests that differences in verb sense may reflect pragmatic relations, rather than lexical knowledge of individual verbs. Finally, comprehension question accuracy (averaged over all target and filler sentences) was high ($M = 92\%$, $SE = 1\%$) and no individual participant had an error rate greater than 20%. Furthermore, there was no difference in error rates between the congruent ($M = 88\%$, $SE = 1\%$) and incongruent ($M = 89\%$, $SE = 1\%$) conditions [ANOVA: both $F_s < 1$; Mixed-effect regression: Wald $Z < 1$].

Discussion

The results of the self-paced reading experiment are consistent with our hypothesis that comprehenders rely on their knowledge of common events to form expectations for upcoming sentence participants. Reading times on the two words immediately following the patient noun were shorter in the congruent condition, where they were consistent with the event evoked by the agent-verb pair. These results could not be due to priming from the verb to the patient noun, because the same verb-patient pairings were used in the congruent and the incongruent conditions. Moreover, as the priming study showed, they are also unlikely to be due to priming from the agent noun.

Thus the reading time results support our claim that comprehenders assess the plausibility of a patient role-filler relative to a specific event at some stage in the comprehension process. Nonetheless, reliable reading time differences were observed starting with the word following the patient noun rather than at the patient itself. There are two possible explanations for this delayed effect. On the one hand, it could indeed be that comprehenders do not make immediate use of event knowledge in sentence processing, and the current effects result from later processes. This would be consistent with findings from the eye movement literature showing only relatively late effects of plausibility during normal reading (e.g., Rayner et al., 2004; Warren & McConnell, 2007).

A second possibility, however, is that comprehenders use their event knowledge immediately, but the self-paced reading time measure failed to capture the timing of these effects. If so, then a more temporally sensitive online measure of incremental sentence processing, such as event-related brain potentials (ERPs), might give more fine-grained timing information. Indeed, numerous ERP studies have shown differential electrical brain activity from 300–500 ms post-item onset for words that are more rather than less congruent with (or differentially expected in) their context. This difference is known as the N400 effect, and is due to smaller amplitudes for expected relative to unexpected words (Kutas & Hillyard, 1980; Kutas & Hillyard, 1984). The N400 effect could thus provide more information about the time course of contextually driven processes in word comprehension. More specifically, to the extent that *mechanic checked* results in stronger expectations for an event that include an automobile (such as *checked the brakes*) than does *journalist checked*, we would expect to see smaller N400 amplitudes to the congruent than incongruent patient.

Experiment 2

To examine the time course of the effect of event knowledge, we conducted an ERP study using the materials from the self-paced reading experiment.

Method

Participants—Twelve students (7 women, 5 men, mean age=20.5, SD=2 years) from UC San Diego participated for course credit. All were native right-handed speakers of English with normal or corrected-to-normal visual acuity.

Materials—The materials used for the ERP experiment were identical to those used in the self-paced reading.

Procedure—Participants were tested in a soundproof, electrically shielded chamber and were instructed to read words presented on a CRT monitor for comprehension. Each pair of sentences was presented using rapid serial visual presentation (RSVP). Each word was presented along with a fixation dot for 200 ms followed by the fixation dot alone for 100 ms, for a total SOA of 300 ms. It is important to note that the presentation rate in this study (unlike in many ERP studies) is actually faster than the mean self-paced reading speed per word, which was approximately 350 ms. This means that participants had no more time to read the words in the ERP experiment than they did in the self-paced reading study; in fact, they had, on average, somewhat less.

There was a 1000 ms break between the two sentences of each pair. Each pair was followed by a yes-no comprehension question. The study was divided into four blocks, and participants were given a break after each block. The reading task took 20–30 minutes to complete.

The EEG was recorded from 26 tin electrodes distributed evenly over the scalp via an elastic cap, referenced to the average of the left and right mastoids. Eye movements and blinks were monitored via electrodes placed on the outer canthus and infraorbital ridge of each eye. Electrode impedances were kept below 5 K Ω . The EEG was amplified with Nicolet amplifiers with a bandpass of .016–100 Hz and was digitized at 250 samples per second.

Analysis—We computed the average ERP to the patient nouns using epochs extending from 500 ms before the onset of the noun to 1548 ms after. Any trials contaminated with blinks, eye movements, excessive muscle activity, or amplifier blocking were rejected offline before averaging. To analyze for differences in N400 amplitude, we entered the mean amplitude between 300–500 ms post-stimulus, relative to a pre-stimulus baseline of 500 ms, for each participant in each condition into a 2×26 repeated measures analysis of variance, with congruence and electrode channel as factors.²

Results

The grand average ERP in response to the patient noun at all recording sites is shown in Figure 2. As predicted, there was a main effect of congruence in the N400 response to the patient noun, such that the incongruent condition was more negative [-1.63 vs. 0.18 μ V; $F(1,11) = 4.97$, $p < .05$]. There was also a main effect of electrode channel [$F(25,275) = 2.20$, $p < .01$] and an interaction between the two [$F(25,275) = 4.67$, $p < .0001$] indicating that the effect was not evenly distributed across the scalp. Visual inspection revealed that it was largest over

²The traditional software used for ERP data analysis makes it difficult to obtain the by-item data that would be needed to perform a by-items ANOVA or a mixed-effects regression. Thus, as is standardly done, we performed only a by-participants ANOVA for the ERP data.

midline and medial channels, especially at central and posterior recording sites – a distribution typical of N400s in the visual modality.³

Discussion

As predicted, our results show a differential electrical brain response beginning at the patient noun in the congruent versus incongruent conditions. More specifically, within 300 ms of the onset of the patient noun, the ERP in the congruent condition becomes positive relative to that for the patient noun in the incongruent condition; this potential difference lasts for 500–800 ms at various channels, more so in some participants than others, in part because of the fast (300 ms/word) rate of presentation. Both the initial timing and scalp distribution (maximal over right parietal sites) of this contextually driven ERP effect (300–500 ms post patient onset) are consistent with those of the canonical written word N400 congruity effect, a well-established measure of semantic expectancy (integration) in this type of sentence reading paradigm (Kutas & Hillyard, 1984). The ERP data show clear evidence of the differential processing consequences of different agent-verb pairings on the same patient nouns within 300 ms of their appearance, a result that contrasts somewhat with the self-paced reading result in which reading time differences showed up on each of the two words following the patient. These results are thus consistent with the proposal that comprehenders, using their knowledge of common events, differentially process (whether in an anticipatory or integrative fashion) upcoming arguments as they read (or hear) sentences.

General Discussion

Previous research has presented convincing evidence that comprehenders are extremely sensitive to the fit of potential fillers to a verb's thematic roles. But what information do comprehenders rely on in determining that fit? One possibility is that this is lexical information, stored in and retrieved from a verb's lexical representation. In this article, we offer an alternative hypothesis, namely, that comprehenders' expectations regarding likely fillers of the patient role involve knowledge of typical events and situations in the world, and that the event interpretations in turn arise from a dynamic computation that involves multiple cues in the sentence, and the discourse context as well, when one is available. Evidence from the self-paced reading and ERP studies reported here support this hypothesis. The stimuli in these studies were developed through a set of norming studies designed to reduce the possibility of independent word-word facilitation from agent to patient nouns, and to increase the possibility that the experimental items reflected the comprehenders' generalized knowledge of events. In a self-paced reading study using these items, we found that reading times in the post-patient region were faster when the patient noun was consistent with the event evoked by the agent-verb combination. An ERP study using these same materials found effects starting at the patient noun, and lasting beyond. Crucially, with the ERP experiment, we found this immediate effect in a reading task, eliminating Rayner et al.'s (2004) objection that the early timecourse of the effects found in studies such as Kamide et al. (2003) depended on the paradigm's artificial limitation of the possible set of role fillers, arguing strongly that rapid plausibility effects like those found in visual world eyetracking studies may be a natural result of the comprehension process itself. We interpret these results as supporting the idea that comprehenders compute expectations based on all available information, including their knowledge of typical events.

However, there are findings which appear to challenge this idea. For example, in a recent eye-tracking study (Rayner et al., 2004), participants read three types of sentences: a set in which the patient was plausible, given the preceding instrument and verb (*John used a knife to chop*

³The difference in means that begins earlier than the 300–500ms window is not reliable: an analysis with a 200–300ms window reveals no effect of congruence [$F < 1$].

the large carrots), a second set in which the verb-instrument combination rendered the patient implausible (*John used an axe to chop the large carrots*), and a third in which the choice of verb made the patient anomalous (*John used a pump to inflate the large carrots*). The authors report an immediate effect of anomaly, in which gaze duration on the patient noun and the two following words was elevated in the anomalous condition, relative to the plausible condition. Effects of plausibility, however, were delayed, with no difference between the plausible and implausible conditions at the patient noun itself in either first fixations or gaze duration. Note that there was some evidence of a relatively early plausibility effect, in that there was a marginally significant plausibility effect at the patient noun in go-past reading times (all fixations from the first in a region until the reader moves past that region, including regressions back to earlier regions). This pattern of results is more consistent with a lexical retrieval account, which could explain the immediate difference between the anomalous and control conditions as a response to a violation of the verb's selectional restrictions (although Rayner et al. did not interpret their results in terms of lexically-based selectional restrictions). The delay in the plausibility effect could be taken to suggest that the evaluation of plausibility requires higher-order, non-lexical information, which takes longer to compute. Similar results have been reported by Warren and McConnell (2007), who did conclude that immediate effects depend on selectional restriction violations.

Matsuki, Chow, Hare, Elman, Scheepers, and McRae (2009), however, presented evidence that instrument-verb-patient plausibility manipulations can in fact produce significant, immediate effects. They performed both a self-paced reading and an eye tracking experiment, using instrument-verb-patient triplets obtained – as in this paper – from norming studies designed to probe comprehenders' knowledge of typical events, and showed the predicted effect of plausibility on the patient noun in self-paced reading, first fixation, and gaze duration measures. There are four possible factors that could have contributed to this difference in results. One possibility, supported by Matsuki et al.'s norming experiment in which they asked participants to, for example, list the types of things that people chop with a knife, is that the difference in typicality between their typical and atypical conditions was larger than that between, for example, Rayner et al.'s (2004) control and implausible conditions. However, these differences were not evident in a plausibility rating study conducted by Matsuki et al. (both Rayner et al. and Warren and McConnell used plausibility ratings to construct their items). A second possibility is that the earlier effects might have been obtained because Matsuki et al.'s items were fully crossed, whereas Rayner et al.'s and Warren and McConnell's were partially crossed (although the patients were the same in all conditions in the latter two studies). Third, Matsuki et al. used 48 items, whereas the other studies used 30, so that the increase in power might be responsible. A final possibility, pointed out by Warren (in press), is that differences in the timing of plausibility effects in eye tracking may result from complex interactions between the time courses of saccade programming, lexical processing, and sentence-level processing, which are as yet not fully understood.

Another question that arises is why we found an apparent difference in the time course of the effect of congruence between the self-paced reading and ERP experiments. Specifically, with ERP there was a significant difference between the congruent and incongruent conditions evident in the N400 response timelocked to the critical patient noun. However, with the self-paced reading results, the difference between the conditions only became significant at the response to the first word after the patient. One simple explanation is that this reflects intrinsic differences between a continuous measure (ERP) and one sampled only every 350 ms or so (self-paced reading time). If, for example, a difference arose about 300 ms after the onset of the patient noun, that may be too late for it to affect the initiation of a motor response that occurs on average every 350 ms, while it would still appear in an electrophysiological measure sampled every 4 ms. Under this explanation, the behavioral measure should show up on the word following the critical word. Other evidence, however, indicates that ERP may at times

be a more sensitive measure, registering differences which are not apparent in self-paced reading data. Ditman, Holcomb, and Kuperberg (2007) presented data from an experiment with a plausibility manipulation using concurrent ERP and self-paced reading methodologies that showed a significant N400 effect in response to the critical word, but in which self-paced reading times showed a reliable effect only on the final word of the sentence, three words later (see also Van Berkum et al., 2005, for a similar difference in time course between ERPs in response to auditory language comprehension and self-paced reading.) Our data may be yet another example of this difference in the sensitivity of these two dependent measures. That said, it remains an open and important question as to what the exact relationship is between the motor commands initiating self-paced reading responses (and eye movements during reading as well) and their relationship to the generators of ERP components such as the N400. Studies such as Ditman et al. (2007), which monitor ERPs concurrently during a self-paced reading task, seem an especially promising way to gain insight into these relationships.

Three accounts

In the experiments presented here, we demonstrated that comprehenders combine knowledge about an agent and a verb to influence their expectations for upcoming patient nouns. There are at least three accounts of the type of knowledge that comprehenders may use to do this. Specifically, comprehenders may use (a) their lexical knowledge, (b) their knowledge of statistical linguistic regularities, or (c) their knowledge of typical events. We discuss each of these possibilities in turn.

It appears at first glance that one could derive a lexical sense-based approach that would account for the present results. For example, the lexicon might include multiple entries for each verb, each corresponding to a verb's sense (Cruse, 1986; Klein & Murphy, 2001; Lehrer, 1990; but see also Caramazza & Grober, 1976; Williams, 1992). For example, *check-1* may correspond to inspecting mechanical objects, whereas *check-2* may correspond to verifying the correctness of writing. Under this account, the agent noun (mechanic or journalist) would activate the most likely sense of the verb. It could further be assumed that each sense includes lexically-based information regarding the likelihood of fillers of the patient role (McRae, Ferretti, & Amyote, 1997). Therefore, each sense would activate conceptual information regarding potential upcoming patient nouns, and thus brakes or spelling would be differentially expected. The major difficulty for such an approach in accounting for the present results is that it is unlikely that there are verb sense differences for the majority of our items, and so lexical information alone would not be sufficient.

The second possibility is that comprehenders are using their knowledge of statistical linguistic regularities. For example, in a class of computational models of word relations known as *topic models* (Blei, Ng, & Jordan, 2003; Griffiths & Steyvers, 2002, 2003, 2004; Griffiths, Steyvers, & Tenenbaum, 2007), words are modeled as providing evidence for the topic underlying a particular discourse, and the inferred likely topic (or distribution over topics) from words already encountered can be used to predict upcoming words. To take an example from our own stimuli, because these topics can be arbitrarily specific, the words *mechanic* and *checked* might activate a different topic than *journalist* and *checked*; in the topic activated by the former, *brakes* may be a likely upcoming word, while in the topic activated by the latter, *spelling* might be relatively more likely.

If these relationships are expressed in language, they might be recoverable by a suitably sophisticated model. Such a model, however, would also need to capture the overall message being expressed, and recent work indicates that this requires an approach that integrates knowledge that does not appear to be present in such topic models. Negation, for example, has rapid effects on the N400 (Nieuwland & Kuperberg, 2008), and the scope of negation varies with the precise arrangement of words. Similarly, Otten and Van Berkum (2007) found that

the N400 effects in discourse are influenced by precise constraints of the message, rather than a scenario suggested by a few of the content words. Hence, although more complex models of word relations such as topic models can capture important aspects of the comprehension process, there is much still to be explained.

This is not to say that linguistic knowledge has no influence independent of event knowledge, or, conversely, that spoken and written language offer a veridical account of the event being described. A speaker may construe the same event in a number of ways, and use their knowledge of language to express that construal. Furthermore, there are cases where aspects of participants in an event are unlikely to be mentioned, as for example when they are in common ground. The knowledge that underlies these examples, however, requires a level of conceptual and linguistic complexity that goes well beyond the tracking of co-occurrence statistics.

Thus, we favor the third account, in which comprehenders immediately and dynamically combine a range of cues including event knowledge in language processing. Under this account, the linguistic material a comprehender encounters activates knowledge of a likely class of events. Specifically, in our experiments, an event was activated by a particular combination of agent and verb, but there is other evidence that this activation may also take discourse cues into account (Race, Klein, Hare, & Tanenhaus, 2009). Once this event is activated, the comprehender's knowledge of typical events facilitates processing of words congruent with the event. We showed facilitation of a likely patient noun, when filling the verb's patient role. Other evidence is suggestive that event knowledge may facilitate words that are congruent with an event, even when they fill the wrong role (Metusalem, Kutas, Hare, McRae, & Elman, 2010). Converging evidence for this interpretation comes from a range of work in the visual world (e.g., Knoeferle & Crocker, 2006, 2007; Altmann & Kamide, 2007, 2009), priming (e.g., Chwilla & Kolk, 2005; Duffy et al., 1989), and other studies of reading (e.g., Nieuwland & van Berkum, 2006; Morris, 1994; Filik, 2008; Bicknell & Rohde, 2009).

Conclusion

We began this article with the goal of showing that comprehenders' knowledge of events in the world plays a major role in language interpretation. The present results are consistent with this claim. As we have shown, the choice of agent noun in a sentence can shade the interpretation of the event that the verb evokes, and this, in turn, influences the processing about the upcoming patient. An account based solely on lexical retrieval would find this result difficult to explain, even if different verb senses encoded separate information about thematic role fillers, because the sense variations in many of our items are too subtle to allow any but the most arbitrary division into separate lexical entries. Instead we argue that comprehenders dynamically compute a context-dependent interpretation based on all available cues, including their knowledge of typical events and situations in the world.

Acknowledgments

We thank Roger Levy, Andy Kehler, Hannah Rohde, Florian Jaeger, Tom Urbach, Jim Magnuson, Tessa Warren, and an anonymous reviewer for help and useful discussion about this work, in addition to the audiences at the 82nd Annual Meeting of the Linguistic Society of America, the 30th Annual Meeting of the Cognitive Science Society, and the 21st Annual CUNY Conference on Sentence Processing. This research was supported by NIH Grants MH60517, HD053136, and HD22614, and NSERC Grant OGP0155704. The first author was supported by the NIH Training Grant 'Language, Communication and the Brain' awarded to the Center for Research in Language at the University of California, San Diego (T32-DC000041).

Appendix A

Stimuli

1. address

(Cong, group) The speaker addressed the group that was assembled at the park. He commended the community on its efforts in rebuilding the city and pledged to support more public works in the future.

(Incong, group) The secretary addressed the group that was assembled at the meeting. She had put together a list of new office supplies and needed the group's approval.

(Cong, letter) The secretary addressed the letter to the president of the company. She had never really liked her boss, and wanted to find a new job.

(Incong, letter) The speaker addressed the letter to the president of France. With the election coming up, he felt he had to make his grievances known.

2. arrange

(Cong, score) The composer arranged the score for a new film directed by Martin Scorsese. He had been handpicked by the renowned director for his lifetime of beautiful work.

(Incong, score) The librarian arranged the score for a new film directed by local high school students. They picked her because they knew she played several musical instruments in her spare time.

(Cong, shelf) The librarian arranged the shelf that displayed some old classics. She hoped that if they were prominently displayed, they might interest some young readers.

(Incong, shelf) The composer arranged the shelf that displayed some old sheet music. He liked to display the music to his favorite songs to continually inspire him.

3. bend

(Cong, back) The contortionist bent her back painfully while attempting to put her leg behind her head. It was a maneuver she had done countless times before, but it seemed her age was finally catching up with her.

(Incong, back) The blacksmith bent his back painfully while attempting to pick up the horseshoe he had dropped. As he got older, it kept getting harder and harder to manage his shop alone.

(Cong, rod) The blacksmith bent the rod that was used for tending fires. Distracted by something else, he had left it in the fire a bit too long.

(Incong, rod) The contortionist bent the rod that was used for supporting her legs during her act. She hadn't realized exactly how much weight she was putting on it, until it finally gave out.

4. blast

(Cong, radio) The teenager blasted the radio so loudly that he woke up his parents downstairs. They were quickly becoming less than enchanted with his new tastes in music.

(Incong, radio) The soldier blasted the radio so loudly that he woke up his commander next door. He was quickly reprimanded and assigned to extra cleaning duties.

(Cong, building) The soldier blasted the building with a powerful bomb that destroyed nearly the whole block. His target was an area away from the center of the town, but there were still many civilian injuries and casualties.

(Incong, building) The teenager blasted the building with a powerful bomb that he made in his garage. He picked one of the city buildings where he knew he could make the most damage.

5. break

(Cong, leg) The skier broke his leg while he was trying to do a 360. It was a move for an expert skier, and he didn't have nearly enough experience to even attempt it.

(Incong, leg) The smoker broke his leg while he was trying to clean his gutters. While he took a nasty fall, the fact that the smoking had weakened his bones made the injury much worse.

(Cong, habit) The smoker broke the habit that had been ruining his lungs. His wife begged him to stop after watching his health deteriorate over the years.

(Incong, habit) The skier broke the habit that had been ruining his jumps. He tended to brace himself out of fear, which slowed him down too much.

6. check

(Cong, spelling) The journalist checked the spelling of his latest report to the editor. His editor was a stickler for correct spelling and made everyone else at the paper obsess about it too.

(Incong, spelling) The mechanic checked the spelling of his latest report about the engine. Not having had much formal schooling, he strived to maintain professionalism at his shop.

(Cong, brakes) The mechanic checked the brakes on the car he had gotten earlier that day. It turned out that the rear passenger-side brake was the source of the complaint.

(Incong, brakes) The journalist checked the brakes on the car he had rented earlier that day. The car had been acting strangely on the way to his interview.

7. comb

(Cong, hair) The performer combed his hair carefully in the mirror before he went on stage. This performance was going to be broadcast, and he wanted to look perfect for his first TV appearance.

(Incong, hair) The detective combed his hair carefully in the mirror before he met his new client. He knew that first impressions were very important, especially in the private investigator business.

(Cong, street) The detective combed the street looking for clues as to what had happened to the victim. She had been snatched so quickly that it seemed likely some of her personal effects would have been left behind.

(Incong, street) The performer combed the street looking for clues as to what had happened to his partner. He hadn't showed up for their act, which was really unlike him.

8. complete

(Cong, form) The patient completed the form that was required by the doctor before they would even look at his leg. He groaned while going through all their formalities while in excruciating pain.

(Incong, form) The author completed the form that was required by the publisher before they would even look at his manuscript. The fact that he was being filed away as just another number made him think his chances were slim.

(Cong, article) The author completed the article just in time for the publisher's deadline. No matter how early he started writing, he never seemed to finish ahead of schedule.

(Incong, article) The patient completed the article just in time for the nurse to call him in from the waiting room. He had been fascinated to learn all about the childhood of Goldie Hawn.

9. dodge

(Cong, punch) The boxer dodged the punch thrown at him by his opponent. He then lunged forward and threw a right hook that took his opponent by surprise.

(Incong, punch) The politician dodged the punch thrown at him by his opponent. It was a spectacle that made the evening news across the state, as the civil debate had turned ugly.

(Cong, truth) The politician dodged the truth whenever he was asked about his past. He knew that if word got out about his younger years as a family planning advocate, his hopes for election would be over in his conservative state.

(Incong, truth) The boxer dodged the truth whenever he was asked about his past. He knew that if word got out about his former use of steroids, he would lose all the titles he had won.

10. evade

(Cong, question) The candidate evaded the question that was asked of him by the skeptical reporter. He knew that if he took the hard stand that the reporter wanted this late in the election, his chances would all be over.

(Incong, question) The fugitive evaded the question that was asked of him by the detective in the interrogation chamber. He didn't want to tell them where he had been the night before, but hadn't yet thought of a good alibi.

(Cong, cops) The fugitive evaded the cops by ducking into an alleyway. The police were hot on his trail after he was seen in a nearby convenience store.

(Incong, cops) The candidate evaded the cops by ducking into an alleyway. The police were hot on his trail after his competition had mysteriously fallen ill.

11. finish

(Cong, story) The writer finished his story about the time that he spent with the Bedouin tribe. He was inspired by their philosophies that challenged the Western point of view, and wanted to share them with everyone.

(Incong, story) The glutton finished his story about the time that he exhausted the restaurant's buffet. His buddies doubled over in laughter, especially when he told them that they had barred him from ever coming back.

(Cong, meal) The glutton finished his meal at his favorite restaurant after three hours of continuous eating. He was pleased to live in a city with so many all-you-can-eat buffets, which were the only type of restaurant that he would go to.

(Incong, meal) The writer finished his meal at his favorite restaurant feeling perfectly contented. It was the fourth time the newspaper had sent him there, and he planned to write up another excellent review.

12. fire

(Cong, rifle) The marine fired the rifle at his ungrateful new commander. He had had enough of his commander's sadistic tendencies to punish the whole unit, and finally had to stand up to him.

(Incong, rifle) The boss fired the rifle at his ungrateful new employee. He had jumped off the deep end when he heard that the new worker had given away all of the company's secrets to the competition.

(Cong, accountant) The boss fired the accountant who had been working to prepare his company's financial report. The accountant had balked when asked to cook the books.

(Incong, accountant) The marine fired the accountant who had been working to prepare them for their biannual audit. The accountant clearly didn't know the first thing about how finance worked in the armed services.

13. fix

(Cong, television) The technician fixed the television by hitting it on its side after he had double-checked the connections and the wiring. He had no idea why that had worked, but he still charged the customer full price.

(Incong, television) The cook fixed the television by hitting it on its side when it went out during the World Cup. In the game's last five minutes, the entire kitchen staff had come out to watch.

(Cong, lunch) The cook fixed his lunch quickly before he rushed out the door to go to work. His new restaurant had an unfortunate policy of not letting their employees eat restaurant food.

(Incong, lunch) The technician fixed his lunch quickly before he rushed out the door to go to his first call of the day. He always brought his lunch with him, because it was hard to predict how long each job would take.

14. grasp

(Cong, saw) The carpenter grasped the saw very awkwardly as he tried to cut the board. Although the salesman promised that the saw's innovative new grip would double his efficiency, he couldn't quite figure out how to use it.

(Incong, saw) The scholar grasped the saw very awkwardly as he tried to cut the wood for his fireplace. Not being much of an outdoorsman, he didn't know the first thing about using handtools.

(Cong, concept) The scholar grasped the concept very easily after his colleague explained it to him. It was a brilliantly simple idea relating population weight gain to world market forces.

(Incong, concept) The carpenter grasped the concept very easily after his foreman demonstrated it to him. If done correctly, one nail could take the place of three in many cases.

15. load

(Cong, truck) The farmer loaded the truck with several large boxes of his latest crop of tomatoes. He was excited to take them to market, because he knew they would fetch top dollar.

(Incong, truck) The assassin loaded the truck with several large boxes of explosives. He had just bought two tons of gunpowder, which he was planning to use to blow up the Governor's estate.

(Cong, pistol) The assassin loaded the pistol in preparation for killing the president. He only had a five second window, and he knew he couldn't fire more than a couple of shots in that time.

(Incong, pistol) The farmer loaded the pistol in preparation for killing the coyote who had been attacking his cows. He could hear it howling loudly in the full moon, and was determined to track him down tonight.

16. name

(Cong, newborn) The mother named her newborn in honor of her great-grandmother. It made her shed a tear as she realized her daughter would never get to meet her namesake.

(Incong, newborn) The suspect named her newborn in honor of her alleged victim. While many took this as further evidence that she could not have killed him, some skeptics thought it was done purely to generate public sympathy.

(Cong, accomplice) The suspect named her accomplice in the murder of her latest victim. After being double-crossed, she didn't feel the slightest bit of remorse in ratting her out.

(Incong, accomplice) The mother named her accomplice in the murder of her husband as part of her plea bargain. The DA had finally found a way out of giving her a life sentence for killing her children's abusive father.

17. operate

(Cong, cartel) The druglord operated the cartel for over twenty years from his home in Caracas. Everyone there knew who he was and where he lived, so it was no surprise when he finally got caught.

(Incong, cartel) The worker operated the cartel for over twenty years before he started working at the tire plant. He had gotten worried that the police would catch up to him and thought an ordinary job would keep him safe.

(Cong, machinery) The worker operated the machinery that was used to carry heavy boxes. He had to be careful not to run in to things with so much momentum.

(Incong, machinery) The druglord operated the machinery that was used to synthesize meth. It wasn't normally his job, but the normal guy had just disappeared.

18. plant

(Cong, camera) The spy planted the camera inside of the large painting hanging in the Senator's living room. This way, he could see every detail of the next day's secret meeting.

(Incong, camera) The gardener planted the camera inside of the large bush in his garden. Someone had been stealing his vegetables, and he was determined to find out who.

(Cong, vegetables) The gardener planted some vegetables around the Senator's house and cared for them regularly. The Senator's wife liked to cook with vegetables from their yard, but didn't know the first thing about caring for plants.

(Incong, vegetables) The spy planted some vegetables around the Senator's house posing as a gardener. He was one of a team of Russians trying to gather information about important people in the US government.

19. pluck

(Cong, eyebrows) The woman plucked her eyebrows before she went out on her date. She spent hours getting ready and didn't want a single hair to be out of place.

(Incong, eyebrows) The gardener plucked her eyebrows before she went out to the convention. When interacting with so many vendors, she wanted to look as professional as possible.

(Cong, rose) The gardener plucked a rose on the way to her boss's house. She thought he might want to see the first bloom of the season.

(Incong, rose) The woman plucked a rose on the way to visit her friend in the hospital. She thought a happy, yellow rose would brighten her day.

20. prepare

(Cong, dinner) The chef prepared the dinner for the meeting that was held at his restaurant. The restaurant had a special room used frequently by the city's businesspeople.

(Incong, dinner) The politician prepared the dinner for the meeting that was held at his house. He took extra care with the meal so that it would give no reason for complaint.

(Cong, speech) The politician prepared the speech that he would give at the upcoming Democratic rally. He made sure to emphasize his family history of military service, as well as his strong commitment to veteran's affairs.

(Incong, speech) The chef prepared the speech that he would give at his retirement party. Looking back on his long career and flipping through old photos brought back many happy memories for him to share.

21. produce

(Cong, puppet) The toymaker produced the puppet that was used as the prototype for the company's new doll all by himself. He wowed the executives with his skillful craftsmanship.

(Incong, puppet) The moviemaker produced the puppet that was used as the villain in the new children's film all by himself. He had a perfect vision of how he wanted it to look, and he wasn't willing to trust anyone else to do it exactly right.

(Cong, blockbuster) The moviemaker produced the blockbuster with the money that he had put aside for a pet project of his own. He felt his idea for the film was excellent, but he could never get the company executives to agree.

(Incong, blockbuster) The toymaker produced the blockbuster with the money that he earned as the CEO of Mattel. The company's latest line of dolls were prominently featured in the movie, and he was willing to bet that it would significantly increase their sales.

22. raise

(Cong, hand) The student raised her hand at the end of class to ask whether the teacher wanted them to hand in the assignment that was due. The other students all groaned, thinking the teacher would not have remembered without her.

(Incong, hand) The parent raised her hand at the end of the PTA meeting. She had a proposal for a new crosswalk location that she wanted the group to discuss.

(Cong, kid) The parent raised her kid as best she could while working two full-time jobs. Despite the time spent away from her son, she knew she was making a great investment in his future.

(Incong, kid) The student raised her kid as best she could while getting her Bachelor's. Being a full-time student and parent was difficult, but she knew it had to be done to get a good job.

23. report

(Cong, friend) The tattletale reported his friend to the investigators after he saw drugs in his backpack. His normally annoying behavior actually proved helpful in the drug bust.

(Incong, friend) The scientist reported his friend to the investigators after it came out that he had falsified his results. He felt bad about betraying his friend, but he knew it was the right thing to do.

(Cong, results) The scientist reported the results of his most recent experiment examining the function of dopamine in the brain. The controversial study demonstrated that the chemical was crucial to mental health.

(Incong, results) The tattletale reported the results of his most recent attempt at spying on his next-door neighbor to his parents. Unfortunately, they weren't as bothered as he was that the teenager next door was crossing the street without an adult.

24. roll

(Cong, joint) The pothead rolled a joint before he started making dinner every day after work. What was once a social routine had quickly become a daily habit he needed to relax.

(Incong, joint) The baker rolled a joint before he started making dough for the next day. He never could get through the late night shift without lighting up at least once.

(Cong, dough) The baker rolled the dough out completely flat to make a crispy pastry. The secret to his impossibly tender pastry crust was rolling the dough out to an eighth of an inch.

(Incong, dough) The pothead rolled the dough out completely flat to make a late-night pizza. Late nights smoke-outs always gave him the munchies.

25. save

(Cong, five dollars) The shopper saved five dollars by buying all of his groceries in bulk. The savings weren't tremendous but it allowed him to stock up on things he knew he would run out of quickly.

(Incong, five dollars) The lifeguard saved five dollars by buying all of his sunscreen in bulk. The savings weren't tremendous, but he knew he would have to go through a lot of it anyway.

(Cong, boy) The lifeguard saved the boy who had gotten lost in the waves. The boy had ventured out too far and had been swept away by the undertow.

(Incong, boy) The shopper saved the boy who had gotten lost in the mall. A mother herself, she was immediately concerned when she saw the young child walking alone.

26. shine

(Cong, light) The policeman shined his light in the attic looking for the perpetrator hiding there. They had tracked him down to this house, but had exhausted all of the possible hiding places on the ground floor.

(Incong, light) The butler shined his light in the attic looking for the old box of photos. His employer had suddenly been overcome with a desire to see his old wedding pictures.

(Cong, shoes) The butler shined his shoes meticulously before going to work every day. He took pride in every aspect of his traditional English attire.

(Incong, shoes) The policeman shined his shoes meticulously before going to work every day. He was rather old-fashioned and tried to keep his uniform looking its best.

27. spread

(Cong, news) The gossip spread the news about her competitor's recent failure. Apparently, their latest product had to be recalled right after it was released.

(Incong, news) The caterer spread the news about her competitor's recent disaster. Their latest job was a wedding banquet where all the guests ended up getting salmonella.

(Cong, butter) The caterer spread the butter very slowly as she was preparing the sandwiches. They were for a very important tea luncheon at the city's country club.

(Incong, butter) The gossip spread the butter very slowly as she detailed the story about her boss's affair. She couldn't wait to tell her coworkers the juicy story, and spent far more time talking than eating.

28. stuff

(Cong, mouth) The child stuffed his mouth full of the chocolate chips that his mother was going to use to make cookies. He snuck into the kitchen when she went to answer the door, hoping she wouldn't notice.

(Incong, mouth) The cook stuffed his mouth full of the chocolate chips that were supposed to be used for a cake. Despite the health codes that specified otherwise, he could never resist eating on the job.

(Cong, turkey) The cook stuffed the turkey with the seasoned breadcrumb mixture that won awards year after year. It was a traditional American classic that he had refined to perfection.

(Incong, turkey) The child stuffed the turkey with the seasoned breadcrumb mixture to help his mother prepare the Thanksgiving meal. He liked to help his mother in the kitchen so much, and she thought it was a relatively easy job to give him.

29. tie

(Cong, match) The player tied the match by scoring a goal at the last minute. His teammates cheered loudly and picked him up and carried him away.

(Incong, match) The sailor tied the match by scoring a goal at the last minute. The big army-navy game had ended in another draw.

(Cong, rope) The sailor tied the rope tightly around the new mast to secure the mainsail. He hoped this new mast would prove to be more sturdy than the last one.

(Incong, rope) The player tied the rope tightly around the new bag of soccer balls after practice. He wanted to make sure to keep the new balls separate from the old ones in the storage room.

30. toss

(Cong, frisbee) The player tossed the frisbee really far for the dog to catch. A playful dog had joined in the game of ultimate frisbee they were playing in the park.

(Incong, frisbee) The housekeeper tossed the frisbee really far for the dog, so it would go away. When no one else was home, it followed her all around the house and made it hard for her to work.

(Cong, leftovers) The housekeeper tossed the leftovers when she realized they were over a week old. It was some take-out the kids had gotten on their way home from school, and she didn't know it had been in the fridge.

(Incong, leftovers) The player tossed the leftovers when she realized they were over a week old. She had been playing so many away games lately that nearly half of her refrigerator had gone bad.

31. treat

(Cong, grandchildren) The grandfather treated his grandchildren to ice cream as a reward for winning the little league game. He had fond memories of doing the same for his son thirty years before.

(Incong, grandchildren) The nurse treated her grandchildren to ice cream as a reward for staying so quiet all day at the hospital. Her daughter had left them with her for the weekend, not realizing that she had to work.

(Cong, victim) The nurse treated the victim for burns too hastily in an effort to get through the hordes of patients in the ER. She was reprimanded by the doctor later for doing such a poor job.

(Incong, victim) The grandfather treated the victim for burns too hastily in a rush to get to his granddaughter's birthday party. Unfortunately, he overlooked a severe puncture among the burns, and the patient ended up suing him for malpractice.

32. twist

(Cong, ankle) The dancer twisted her ankle in an attempt to complete a tricky jump. She landed on the corner of her foot, and ended up falling to the floor.

(Incong, ankle) The electrician twisted his ankle in an attempt to reach the wiring in the corner. The wiring was in a hard to reach spot, and he made the mistake of trying to do his work without moving the furniture.

(Cong, cable) The electrician twisted the cable around its base to get it out of the walkway. He didn't want the residents to trip on the wires he took care to install.

(Incong, cable) The dancer twisted the cable around its base to get it off of the dance floor. She wanted to make sure no one tripped on it during the performance.

Appendix B

Details of Agent-Patient Priming Study

Method

Participants—Forty-four UCSD undergraduates participated in the lexical decision experiment for course credit.

Materials—The experiment included 88 congruent agent-patient pairs (e.g., *journalist-spelling* or *mechanic-brakes*). Note that this experiment was conducted prior to the step of constructing sentences for Experiments 1 and 2. During that step, we excluded 12 verbs (and thus 24 agent-patient pairs). Therefore, we describe the Method including all 88 pairs, but report analyses for only the 64 pairs that were used in Experiments 1 and 2. The unrelated condition consisted of the incongruent agent-patient pairs (*mechanic-spelling* or *journalist-brakes*).

Also included were the 32 semantically-similar item pairs from McRae and Boisvert (1998). The related prime-target pairs were semantically similar object concepts (e.g., *coconut-pineapple* or *closet-dresser*). These items were included to ensure that priming could indeed be obtained with the participants and procedure that we used (these items produced facilitation in that study). The unrelated condition was created by re-pairing the related primes and targets, as in *closet-pineapple*.

Two lists were created, with each containing half of the items in each set in the related condition, the other half in the unrelated condition. Thus each list contained 44 related and 44 unrelated agent-patient pairs, and 16 related and 16 unrelated McRae and Boisvert (1998) pairs. To create a relatedness proportion of .25, 120 unrelated word-word filler pairs were included, for a total of 240 word-word pairs. An additional 240 word-nonword pairs were included so that the correct response for the lexical decision task was *yes* on 50% of the trials.

Two pseudo-randomizations of each list were created, such that the trials involving test pairs (whether agent-patient or McRae and Boisvert, 1998, items) were always separated by at least one trial. In each pseudo-randomization, the McRae and Boisvert items occurred after all agent-patient items had been presented.

Procedure—The experiment was conducted on a Macintosh computer using PsyScope. Participants pushed one of two buttons on a CMU button box to indicate whether each target was a word or non-word. The button box recorded decision latencies with millisecond accuracy. Each participant used the index finger of their dominant hand to indicate that the target was a word, and the other index finger for a nonword response.

Each trial began with a fixation cross for 250 ms, followed by the prime for 200 ms, 50 ms of blank screen (for a total SOA of 250 ms), and finally the target, which was displayed until the participant responded. The inter-trial interval was 1500 ms. Participants were instructed to read and pay attention to the first word, then respond to the second word as quickly and accurately as possible. Participants were given a break after every 40 trials, and the first item after a break was always a filler. An entire session lasted about 30 minutes.

Design—Separate analyses were performed for the agent-patient and the McRae and Boisvert (1998) pairs. For each dataset, we performed analyses of variance, testing for an effect of relatedness on the decision latencies and (arcsine-square-root transformed) error rates. List was included as a between-participants dummy variable and item rotation group as a between-items

dummy variable to stabilize variance that may have resulted from rotating participants and items over the two lists (Pollatsek & Well, 1995).

Because of recent concerns in the literature about the appropriateness of analyses of variance in psycholinguistics (Baayen, 2004; Baayen, Davidson, & Bates, 2008; Dixon, 2008; Jaeger, 2008), we also performed generalized linear mixed-effect regression (Breslow & Clayton, 1993). Mixed-effect regression is a generalization of standard multiple regression that, in addition to the standard fixed effect variables used in multiple regression, allows for the inclusion of random effect variables, such as participants and items. Thus, unlike analyses of variance, in which separate analyses must be performed over participant and item means (i.e., F_1 and F_2), both participant and item effects are included in a single model.

We fit a generalized linear mixed-effects regression with a logit link function (Agresti, 2002, Bates & Sarkar, 2008, Breslow & Clayton, 1993) to the error data, including relatedness as a fixed effect. We also included the logarithm of the target word's frequency in the British National Corpus as a fixed effect, along with random participant and item intercepts and random participant and item slopes of relatedness. (The model also includes a parameter for the correlation between the random intercept and slope variances.) For the response latencies, we fit a linear mixed-effects regression (Pinheiro & Bates, 2000) with the same fixed and random effects as the error model.

For these and all mixed-effect models reported in this paper, all random effects that we describe were included in the initial model, and iteratively removed if they did not significantly increase the model's log-likelihood. Following that, we removed fixed effect covariates which were not significantly different than zero. To assess the significance of a fixed effect in the generalized linear models, we report the Wald Z statistic for the effect's coefficient and its p -value. To assess the significance of a fixed effect in the linear models, we report the t -value of its coefficient, but not the t -value's degrees of freedom, because there is no general agreement on the correct degrees of freedom to use. (A general guideline, however, for large datasets such as ours is that a t -value greater than 2 indicates significance.) For this reason, as is standardly done, we report a p -value obtained from the posterior distribution on the coefficient produced after 50,000 iterations of Markov chain Monte Carlo (MCMC) sampling on the model (Baayen et al., 2008). To avoid collinearity in our models, we centered each of our fixed effects, so that the correlations between fixed effect coefficient estimates were never greater than 0.1

Results & Discussion

Response latencies greater than three standard deviations above or below the grand mean were replaced with the cutoff value, which affected 2.4% of the data for the McRae and Boisvert (1998) items and 1.6% of the data for the agent-patient pairs. Trials on which there were errors were excluded from the decision latency analyses. Decision latencies and percent errors are presented in Table 1. Mean decision latencies for the congruent and incongruent agent-patient pairs were virtually identical [ANOVA: both F 's < 1; Mixed effect regression: $t = 0.50$, $p = .6$]. However, McRae and Boisvert's items showed a significant 18 ms of facilitation [ANOVA: $F_1(1,42) = 8.67$, $p < .006$; $F_2(1,30) = 4.31$, $p < .05$; Mixed effect regression: $t = 2.05$, $p < .05$]. There were no effects of relatedness on error rate, for either the agent-patient pairs [ANOVA: $F_1(1,42) = 1.44$, $p > .2$; $F_2 < 1$; Mixed effect regression: Wald $Z = 0.76$, $p > .3$] or the McRae and Boisvert pairs [ANOVA: $F_1(1,42) = 1.64$, $p > .2$; $F_2 < 1$; Mixed effect regression: Wald $Z = 0.32$, $p > .7$]. Therefore, there was no evidence that the agents independently facilitated their corresponding patients. This argues that any effects found in Experiments 1 and 2 are unlikely to be due to that source.

References

- Agresti, A. *Categorical data analysis*. Second ed. New York, NY: John Wiley & Sons; 2002.
- Altmann GTM, Kamide Y. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 1999;73:247–264. [PubMed: 10585516]
- Altmann GTM, Kamide Y. The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language* 2007;57:502–518.
- Altmann GTM, Kamide Y. Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition* 2009;111:55–71. [PubMed: 19193366]
- Baayen R. Statistics in psycholinguistics: A critique of some current gold standards. *Mental Lexicon Working Papers* 2004;1:1–45.
- Baayen R, Davidson D, Bates D. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 2008;59:390–412.
- Bates, DM.; Sarkar, D. lme4: Linear mixed-effects models using S4 classes. 2008. R package version 0.999375-23.
- Bicknell, K.; Rohde, H. Dynamic integration of pragmatic expectations and real-world event knowledge in syntactic ambiguity resolution. In: Taatgen, NA.; Rijn, H., editors. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*; Austin, TX: Cognitive Science Society; 2009. p. 1216-1221.
- Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 2003;3:993–1022.
- Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 1993;88:9–25.
- Caramazza, A.; Grober, E. Polysemy and the structure of the subjective lexicon. In: Rameh, C., editor. *Georgetown University roundtable on languages and linguistics. Semantics: Theory and application*. Washington, DC: Georgetown University Press; 1976. p. 181-206.
- Chwilla DJ, Kolk HHJ. Accessing world knowledge: Evidence from N400 and reaction time priming. *Cognitive Brain Research* 2005;25:589–606. [PubMed: 16202570]
- Cohen J, MacWhinney B, Flatt M, Provost J. PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, and Computers* 1993;25:257–271.
- Cruse, DA. *Lexical semantics*. Cambridge: Cambridge University Press; 1986.
- DeLong KA, Urbach TP, Kutas M. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience* 2005;8:1117–1121.
- Ditman T, Holcomb PJ, Kuperberg GR. An investigation of concurrent ERP and self-paced reading methodologies. *Psychophysiology* 2007;44:927–935. [PubMed: 17850242]
- Dixon P. Models of accuracy in repeated-measures designs. *Journal of Memory and Language* 2008;59:447–456.
- Duffy SA, Henderson JM, Morris RK. Semantic facilitation of lexical access during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 1989;15:791–801.
- Federmeier KD, Kutas M. A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language* 1999;41:469–495.
- Federmeier KD, McLennan DB, De Ochoa E, Kutas M. The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: an ERP study. *Psychophysiology* 2002;39:133–146. Erratum in: *Psychophysiology*, 39, 684–687. [PubMed: 12212662]
- Filik R. Contextual override of pragmatic anomalies: Evidence from eye movements. *Cognition* 2008;106:1038–1046. [PubMed: 17524387]
- Griffiths, TL.; Steyvers, M. A probabilistic approach to semantic representation. In: Gray, WD.; Schunn, C., editors. *Proceedings of the 24th Annual Conference of the Cognitive Science Society*; Mahwah, NJ: Erlbaum; 2002. p. 381-386.

- Griffiths, TL.; Steyvers, M. Prediction and semantic association. In: Becker, S.; Thrun, S.; Obermayer, K., editors. *Advances in Neural Information Processing Systems*. Vol. 15. Cambridge, MA: MIT Press; 2003. p. 11-18.
- Griffiths TL, Steyvers M. Finding scientific topics. *Proceedings of the National Academy of Sciences* 2004;101:5228–5235.
- Griffiths TL, Steyvers M, Tenenbaum JB. Topics in semantic representation. *Psychological Review* 2007;114:211–244. [PubMed: 17500626]
- Hare M, Jones M, Thomson C, Kelly S, McRae K. Activating event knowledge. *Cognition* 2009;111:151–167. [PubMed: 19298961]
- Jackendoff, R. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press; 2002.
- Jaeger TF. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 2008;59:434–446. [PubMed: 19884961]
- Just MA, Carpenter PA, Woolley JD. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 1982;111:228–238. [PubMed: 6213735]
- Kamide Y, Altmann GTM, Haywood SL. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language* 2003;49:133–156.
- Klein DE, Murphy G. The representation of polysemous words. *Journal of Memory and Language* 2001;45:259–282.
- Knoeferle P, Crocker MW. The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science* 2006;30:481–529.
- Knoeferle P, Crocker MW. The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language* 2007;57:519–543.
- Kutas M, Hillyard SA. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* 1980;207:203–205. [PubMed: 7350657]
- Kutas M, Hillyard SA. Brain potentials reflect word expectancy and semantic association during reading. *Nature* 1984;307:161–163. [PubMed: 6690995]
- Kutas, M.; Lindamood, TE.; Hillyard, SA. Word expectancy and event-related brain potentials during sentence processing. In: Kornblum, S.; Requin, J., editors. *Preparatory States and Processes*. Hillsdale, New Jersey: Lawrence Erlbaum; 1984. p. 217-237.
- Lehrer A. Polysemy, conventionality, and the structure of the lexicon. *Cognitive Linguistics* 1990;1:207–246.
- Matsuki, K.; Chow, T.; Hare, M.; Elman, JL.; Scheepers, C.; McRae, K. Dynamic combination of concepts during on-line language comprehension. (submitted).
- McNamara TP. Theories of Priming: I. Associative Distance and Lag. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 1992;18:1173–1190.
- McNamara TP. Priming and theories of memory: A reply to Ratcliff and McKoon. *Psychological Review* 1994;101:185–187.
- McRae K, Boisvert S. Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 1998;24:558–572.
- McRae K, Ferretti TR, Amyote L. Thematic roles as verb-specific concepts. *Language & Cognitive Processes* 1997;12:137–176.
- McRae K, Spivey-Knowlton MJ, Tanenhaus MK. Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language* 1998;38:283–312.
- Metusalem, R.; Kutas, M.; Hare, M.; McRae, K.; Elman, JL. Generalized event knowledge activated during online language comprehension. In: Ohlsson, S.; Catrambone, R., editors. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*; Austin, TX: Cognitive Science Society; 2010. (pp. XXX).
- Morris RK. Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 1994;20:92–103.
- Nieuwland MS, Kuperberg GR. When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science* 2008;19:1213–1218. [PubMed: 19121125]

- Nieuwland MS, Van Berkum JJA. When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience* 2006;18:1098–1111. [PubMed: 16839284]
- Otten M, Van Berkum JJA. What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Research* 2007;1153:166–177. [PubMed: 17466281]
- Pinheiro, JC.; Bates, DM. *Mixed effects models in S and S-Plus*. New York: Springer Verlag; 2000.
- Pollatsek A, Well AD. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 1995;21:785–794.
- Race, D.; Klein, N.; Hare, M.; Tanenhaus, M. What does a shopper expect to save? The role of event knowledge in verb interpretation. In: Taatgen, NA.; Rijn, H., editors. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*; Austin, TX: Cognitive Science Society; 2009. p. 2219
- Rayner K, Warren T, Juhasz BJ, Liversedge SP. The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2004;30:1290–1301.
- Trueswell JC, Tanenhaus MK, Garnsey SM. Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language* 1994;33:285–318.
- Van Berkum JJA, Brown CM, Zwitserlood P, Kooijman V, Hagoort P. Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2005;31:443–467.
- Warren, T. The influence of plausibility and anomaly on eye movements in reading. In: Gilchrist, I.; Everling, S.; Liversedge, S., editors. *The Oxford Handbook on Eye Movements*. (in press.)
- Warren T, McConnell K. Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin & Review* 2007;14:770–775. [PubMed: 17972747]
- Wicha NY, Bates EA, Moreno EM, Kutas M. Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters* 2003;346:165–168. [PubMed: 12853110]
- Williams JN. Processing polysemous words in context: Evidence for interrelated meanings. *Journal of Psycholinguistic Research* 1992;21:193–218.

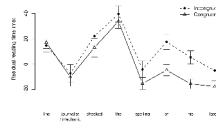


Figure 1. Mean residual reading times. Error bars show one standard error above and below the mean, calculated by participants.

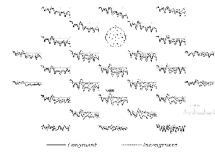


Figure 2. Grand average ERPs at all 26 recording sites (locations marked on iconic head) in response to the patient noun (word 5; onset = 0 ms), in the congruent versus incongruent conditions.

Table 1

Decision latencies (ms) and percent errors for agent-patient priming

	Agent-Patient		McRae and Boisvert	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Decision Latency				
Unrelated	509	9	549	11
Related	509	9	513	12
Facilitation	0		18*	
Percent Errors				
Unrelated	4.8	0.8	11.2	1.4
Related	3.5	0.6	9.5	1.5
Facilitation	1.3		1.7	

* Significant by participants and items

Table 2

Residual and raw reading times (ms) for Experiment 1, self-paced reading

Condition	Region				
	checked	the	spelling	on	his
Residual Reading Times					
Incongruent	22 (7)	36 (5)	-8 (6)	19 (6)	6 (5)
Congruent	20 (7)	35 (4)	-16 (4)	-1 (4)	-12 (3)
Difference	2	1	8	20*	18*
Raw Reading Times					
Incongruent	392 (15)	372 (12)	363 (16)	366 (14)	351 (13)
Congruent	393 (15)	371 (10)	349 (13)	342 (12)	330 (11)
Difference	-1	1	14	24*	21*

Note: Standard errors in parentheses.

* Significant.