Published in final edited form as:

Lang Cogn Neurosci. 2016 ; 31(4): 536–548. doi:10.1080/23273798.2015.1117117.

World knowledge affects prediction as quickly as selectional restrictions: Evidence from the visual world paradigm

Evelyn Milburn, Tessa Warren, and Michael Walsh Dickey

Abstract

There has been considerable debate regarding the question of whether linguistic knowledge and world knowledge are separable and used differently during processing or not (Hagoort, Hald, Bastiaansen, & Petersson, 2004; Matsuki et al., 2011; Paczynski & Kuperberg, 2012; Warren & McConnell, 2007; Warren, McConnell, & Rayner, 2008). Previous investigations into this question have provided mixed evidence as to whether violations of selectional restrictions are detected earlier than violations of world knowledge. We report a visual-world eye-tracking study comparing the timing of facilitation contributed by selectional restrictions versus world knowledge. College-aged adults ($n=36$) viewed photographs of natural scenes while listening to sentences. Participants anticipated upcoming direct objects similarly regardless of whether facilitation was provided by only world knowledge or a combination of selectional restrictions and world knowledge. These results suggest that selectional restrictions are not available earlier in comprehension than world knowledge.

Keywords

language comprehension; plausibility; eye tracking; sentence processing; prediction

Introduction

There is debate as to whether or not there exists specifically linguistic knowledge that can drive language comprehension separately from world knowledge. Classic linguistic theory (e.g. (Chomsky, 1965; Katz & Fodor, 1963)) assumes that there are separate stores for words (the lexicon) and world knowledge, and that only certain semantic properties function within the lexicon. However, this theory has been challenged on multiple grounds. For example, there is debate as to whether lexicons exist (Clark, 1983), as well as whether dividing semantic space into properties internal vs. external to a lexicon is possible (Jackendoff, 2002).

One of the paradigmatic constructs of this classic linguistic theory is selectional restrictions. According to Chomsky (1965), selectional restrictions are lexicon-internal constraints that verbs place on their arguments; they take the form of requirements for specific grammatical features like animacy or being human. For example, under Chomsky's account the fact that the verb *drink* requires an animate agent would be a selectional restriction. Katz and Fodor

(1963) assume that the lexicon incorporates a wider range of semantic features than Chomsky does. Under their account, the fact that *drink* requires a liquid patient would also be a selectional restriction. According to a modular processing theory built on the classically hypothesized representations (e.g., Fodor, 1983), the lexicon is assumed to be a module and therefore its information should be available earlier in the course of language comprehension than information from a non-modular general knowledge base. This account predicts that selectional restrictions should influence processing before world knowledge does. This prediction has been tested in a number of psycholinguistic experiments, most of which have contrasted comprehenders' reactions to violations of selectional restrictions with their reactions to violations of world knowledge (Hagoort et al., 2004; Marslen-Wilson, Brown, & Tyler, 1988; Warren & McConnell, 2007). These studies have been informative, but their evidence is mixed regarding whether violations of selectional restrictions are detected earlier than violations of world knowledge, as will be discussed below. In the experiment reported in this paper, we take the alternative strategy of looking for evidence of processing facilitation associated with selectional restrictions, specifically testing whether such facilitation appears earlier than facilitation due to world knowledge.

In a 2004 paper, Hagoort et al. reported ERP and fMRI experiments designed to determine whether selectional restriction violations are detected earlier than violations of world knowledge. They compared brain responses to sentences that were natural e.g. "Dutch trains are yellow", sentences that violated their participants' world knowledge e.g. "Dutch trains are white", and sentences in which there was a mismatch between the semantic features of a noun and predicate adjective e.g. "Dutch trains are sour". They found no differences in the latency of N400 responses to the two violation conditions, but the violation conditions showed different patterns of oscillations in gamma and theta frequency bands. Hagoort et al. interpreted these findings as suggesting that violations generated by mismatches between lexical semantic features were detected no earlier than violations of world knowledge. However, Pykkänen, Oliveri, and Smart (2009) argued that Hagoort et al. (2004)'s semantic violations were not true mismatches that blocked semantic composition in a linguistic representation, but rather a different style of world knowledge violation. When Pykkänen et al. used MEG to test participants' sensitivity to verbal *un*-prefixation, which they argued implemented a true semantic mismatch, they found differences in the brain structures supporting linguistic versus world knowledge violations.

One concern about these neurolinguistic studies is that their stimulus sets include high proportions of anomalies, and this might influence readers' approaches to the reading task. In an independent set of eye-tracking studies, Warren and colleagues kept the proportion of anomalous sentences low. Additionally, they completed extensive norming to ensure that the events described by their selectional restriction violation (which they operationalized as a mismatch between the basic semantic features a verb requires of its object, and the features of the encountered object, following Katz and Fodor (1963)) and world knowledge violation conditions were as comparable as possible along various dimensions. In a 2007 study, Warren and McConnell minimized differences in the likelihood of the events described by selectional restriction violations and world knowledge violations, and found that readers' eye movements still showed earlier and greater disruption to the selectional restriction violations than to the world knowledge violations. More recently, Warren, Milburn, Patson, and Dickey

(2015) tested items in which the events described by the world knowledge violations and selectional restriction violations were similarly impossible, and found early disruption for only the selectional restriction violations. These findings are the clearest evidence so far that selectional restrictions may be processed earlier than world knowledge.

Still, this body of work testing violations remains open to the concern that it may be impossible to perfectly control for violation severity. Although Warren and colleagues went to great lengths to design items that equated event likelihood and possibility, selectional restriction violations were always rated as having slightly lower likelihood and possibility than world knowledge violations. This is a concern because if greater violation severity generates more disruption and there is some noise in exactly when violation detection occurs, more severe violations could show effects in earlier processing measures, even if the violations were detected at the same time on average. This is because a small amount of large disruption will be more likely to generate a reliable effect than a small amount of weaker disruption in leading-edge processing measures. Therefore, a better way to test whether selectional restrictions are available and used earlier than world knowledge would be to test for very early processing facilitation associated with these different kinds of knowledge. This is the aim of the current experiment.

An existing body of work, primarily by McRae and colleagues, has taken a similar approach to this one. McRae and colleagues argue that there is no specifically linguistic information that can influence language comprehension in a manner dissociable from world knowledge. In an initial line of work supporting this contention, they showed that events and event participants prime each other in ways that would not be expected if the representations underlying the priming were verb-related thematic roles, but would be expected if the underlying representations involved world knowledge about events (Ferretti, Kutas, & McRae, 2007; Ferretti, McRae, & Hatherell, 2001; McRae, Hare, Elman, & Ferretti, 2005). In a further line of work, McRae and colleagues have made the argument that effects of world knowledge appear as early in the processing record as possible, or as early as effects of selectional restriction violations have been observed in other studies, and therefore the use of world knowledge during comprehension cannot be delayed (McRae & Matsuki, 2009). For example, Matsuki et al. (2011) found that world knowledge about event participants facilitated processing during the first fixation on a critical word (which is the measure in which Warren & McConnell (2007) observed the earliest selectional restriction violation effects), even when predictability and lexical association were controlled. These findings about the early timing of world knowledge effects are important and suggestive. However, interpreting them as eliminating the possibility that world knowledge is delayed during comprehension requires assuming that either: (1) there is an objective earliest measure in which a particular effect can appear, or (2) the timing of effects associated with a particular manipulation is always identical across studies. Given that neither of these assumptions is unassailable, a stronger test of this hypothesis would be to compare facilitation associated with world knowledge and selectional restrictions within the same experiment. The current experiment does this.

Current Study

The current study used the visual world paradigm, in which participants hear sentences while their eyes are tracked as they look at pictures. Studies using this paradigm have demonstrated facilitative effects of selectional restrictions (Altmann & Kamide, 1999), world knowledge (Chambers, Tanenhaus, & Magnuson, 2004; Kamide, Altmann, & Haywood, 2003), and lexical associations (Borovsky, Elman, & Fernald, 2012; Kamide et al., 2003; Kukona, Fang, Aicher, Chen, & Magnuson, 2011). However, such studies have not directly compared facilitation associated with selectional restrictions and world knowledge, likely because disentangling these two sources of information is challenging given that any stimulus that implements selectional restrictions also creates an event processed with reference to world knowledge. The current study investigated the timing of processing of selectional restrictions and world knowledge by testing whether adding selectional restrictions to world knowledge constraints boosts or speeds anticipatory looks to a target object. If selectional restrictions are language-internal and have priority during language processing, and therefore are available to act as a predictive filter before world knowledge is used, then they should boost prediction compared to conditions in which only world knowledge drives prediction.

There were two sets of experimental items: event-constrained items in which only world knowledge drove prediction, and verb-constrained items in which both world knowledge and selectional restrictions drove prediction. Each experimental condition was compared to its own baseline that had no predictive information (Table 1). To create world-knowledge-based predictions in the event-constrained condition, we paired sentences containing verbs that place few semantic constraints on their direct objects (e.g. *lick*) with pictures depicting a scene onto which the viewer could naturally project an event described by that verb. For example, a picture of a person writing a letter paired with the sentence “Someone will lick the _____” together predict the target word “envelope” (Figure 1a). In this condition, the sentence's agent and verb provide minimal semantic constraint, so predictions must be guided by event knowledge activated by the combination of the scene and the low-constraint verb. In the verb-constrained condition, scenes were paired with verbs that place strong semantic constraints on their potential objects—for example, the sentence “Someone will pop the _____” is highly predictive of “balloon” even without an image (Figure 1b). Including an image with a balloon then contributed additional constraining event knowledge. Each experimental item was paired with a control item created by changing the verb to one that minimally constrains its direct object, and is thus non-predictive even when presented concurrently with an image.

The sentence stimuli in the present study used the semantically minimal agent “someone” to maximize predictive effects of verb selectional restrictions alone and minimize combinatory effects of the agent and verb. In order to imbue our visual stimuli with rich situational information to support event knowledge, we used photographs instead of the clip art figures that are usually used in visual world studies (Staub, Abbott, & Bogartz, 2012).

Methods

Stimuli

There were 32 critical images, half of which were paired with event-constrained/control sentences and the other half with verb-constrained/control sentences. (see Figures 1a and 1b for examples). Nine of the experimental images were obtained from Staub et al. (2012), and the remaining 23 were either drawn from flickr's pool of Creative Commons-licensed images (<https://www.flickr.com/creativecommons/>) or staged by the investigators. All sentences consisted of a semantically empty subject ("someone"), a future-tense transitive verb, and a direct object. All control sentences contained a verb that was not predictive of a specific object given the agents and objects in the scene.

We also created 64 filler stimuli, each of which consisted of a picture paired with a single sentence. Although all fillers used semantically empty subjects, they varied both in post-verbal structure and in their relationship to the accompanying scene. Filler images were either obtained from Staub et al. (2012) or were Creative Commons-licensed images from flickr. All images in the experiment were resized to 1024×768 pixels.

Norming

All stimuli were normed to verify that they were appropriately constrained or unconstrained. Students from the University of Pittsburgh participated for course credit. All were 18 years of age or older and native speakers of English. One of the verbs in the event-constrained condition turned out to be too constrained, so we removed that item from analysis and also removed the verb-constrained stimulus that normed least well to balance our lists. The norming results presented below and all subsequent analyses are therefore over 30 items.

Sentence-only cloze norm—18 participants provided an appropriate word to complete the 60 experimental and control sentences truncated before the target object (e.g. "Someone will fling the ____"). Sentences were presented in pseudo-random order, with no more than two items from the same condition occurring sequentially. This norm was intended to determine the degree to which each verb constrained its direct object; we therefore coded responses by identifying the most common completion for each item rather than whether or not participants provided the experimental target. This enabled us to ensure that our unconstrained sentences and our event-constrained sentences were similarly unconstraining of their objects when presented without context, regardless of whether the most likely object was the target or a different word. Due to experimenter error, two unconstrained verbs, "pick up" and "point at", were not included in the sentence-only cloze norm. These verbs scored 0.1 and 0.3, respectively, for constrained responses in the picture norm. Because response constraint increased for all conditions in the picture norm compared to the sentence-only cloze norm, and picture norm scores for these verbs were very low, we feel confident in assuming that these verbs would have scored even lower had they been included in the sentence-only cloze norm.

When coding, we consolidated responses that referred to the same referent or were unlikely to be differentiated in the naturalistic scene. For example, for the item "Someone will erase

the ____” we consolidated the answers “board,” “chalkboard,” and “blackboard.” The target was the most common completion for 14 of the 15 verb-constrained items, 2 of the 15 event-constrained items, and none of the items in the control conditions. Table 2 shows idealized and actual mean constrained response proportions in the cloze norm.

Response data were analyzed using linear mixed effects models (Baayen, 2008) in the R statistical computing package (R Development Core Team, 2013; ver. 3.0.1) and using the lme4 package (Bates, 2005; ver. 1.1-7). P-values were obtained using the lmerTest package (ver. 2.0-20), except when logit models, which provide p-values, were used. Models included fixed effects of information type (event vs. verb) and constraint (constrained vs. unconstrained). Random intercepts of participant and item were also added to the model. No models containing any random slopes converged. Significant main effects of information type ($\beta=0.77$; $SE=0.14$; $p<.05$) and constraint ($\beta=1.44$; $SE=0.16$; $p<.05$) were qualified by a significant interaction between the two variables. Responses were most constrained in the verb-constrained condition ($\beta=1.47$; $SE=0.31$; $p<.05$). Critically, this interaction confirmed that linguistic stimuli in the verb-constrained condition more strongly constrained their objects than did stimuli in any other condition.

Sentence & Picture cloze norm—A second round of norms was completed by a different set of 20 participants. These norms were identical in procedure to the first round except that each sentence was accompanied by its corresponding scene, presented using a PowerPoint slideshow. Participants were asked to complete each sentence based on the actors and objects they saw in the image. For this round of norms, responses were coded based on whether the participant provided the target object for each stimulus. Data were analyzed using a linear mixed effects model containing fixed effects of information type and constraint, random intercepts of participant and item, and random slopes of information type and constraint within participants; a model that also included the random slope of constraint within items did not converge. Table 3 shows idealized and actual mean target responses in the sentence+picture norm. Participants provided more target responses in the two constrained conditions than in the control conditions ($\beta=4.36$; $SE=0.37$; $p<.05$). There was also a marginal effect of information type ($\beta=1.15$; $SE=0.31$; $p=.06$), and a significant interaction between information type and constraint such that participants provided the target most often in the verb-constrained condition ($\beta=1.62$; $SE=0.71$; $p<.05$). The significant effect of constraint indicates that participants were more likely to respond with the target in the constrained conditions than in the control conditions when both the linguistic and picture stimuli were presented concurrently. The interaction of information type and constraint suggests that the verb-constrained condition provided more potentially prediction-driving information than the event-constrained condition. This is not surprising because both event-related and selectional-restriction information are available in the verb-constrained condition.

Picture-only cloze norm—A final round of norms was completed by participants ($n=14$) who had not completed either of the first two rounds of norming. This norm was intended to measure how much non-linguistic semantic information (event information) was present in just the picture stimulus, with no cues from the accompanying linguistic stimulus.

Participants viewed each image on a PowerPoint slideshow and completed sentences of the form “Someone will ____” based on what they thought would happen next in the image. Because of the open-ended nature of this task, participants provided a wide variety of responses. We were therefore unable to code responses based solely on whether or not the sentence was completed using the target, and instead coded responses based whether the participant provided the target word, a reasonable synonym, or a semantically-related event when completing the sentence. For example, for a picture of a patient in a dentist's chair with the target word “tooth”, the completions “drill a tooth”, “get their teeth cleaned”, and “have cavities” would all be coded as a target response. Results were analyzed using a linear mixed effects model containing the fixed effect of information type and random effects of participants and items; a model that also included a random slope of information type within participants did not converge. There was no significant difference in target responses across information types ($\beta=0.22$; $SE=0.79$; $p=0.78$), indicating that the pictures accompanying event-constrained and verb-constrained sentences contained the same amount of non-linguistic semantic information that could be used to make predictions about direct objects. Table 4 shows target response proportions in the picture-only norm.

The verbs in the different conditions did not differ in their frequency. ANOVAs indicated that raw verb frequencies from CELEX (Baayen, Piepenbrock, & Gulikers, 1995) showed no reliable effects of information type ($F(1, 14) = .62$; $p=.45$), constraint ($F(1, 14) = 2.77$; $p=.12$), nor interaction ($F(1, 14) = .01$; $p=.91$).

All audio stimuli were recorded by a female native speaker of English. Table 5 shows mean verb and determiner durations for each of the four conditions. For verb duration, there was a marginally significant effect of information type ($F(1, 14) = 4.53$; $p=.052$), but no effect of constraint ($F(1, 14) = .002$; $p=.97$), or interaction ($F(1, 14) = .08$; $p=.78$). The verbs in the verb conditions were, on average, 40 ms longer than the verbs in the event conditions (Table 5). If identification is slower for longer words, this could mean that participants may have had less time for anticipatory looks in the verb conditions. We will return to this concern in the Results section. Determiner duration did not vary reliably by information type ($F(1, 14) = .97$; $p=.34$), or constraint ($F(1, 14) = 2.76$; $p=.12$), and there was no interaction ($F(1, 14) = 2.94$; $p=.11$).

Each scene had one target object, and an interest area was drawn around it approximately one degree of visual angle outside of its borders. Because we used naturalistic scenes rather than composed clip-art images, target objects in the image varied in size and location. In order to ensure that this variation was comparable across conditions, we calculated the distance in pixels from the center of the screen (where participants' gaze was directed at the start of each trial) to the approximate center of each target object. The average pixel-distance of the target from the center of the screen did not differ for the two sets of scenes ($t(28) = .364$; $p=.36$): the average distance to the center of the target was 262.4 pixels in the event-information scenes and 246.7 pixels in the verb-information scenes.

It was important that in the control conditions there always be multiple items in the photograph that could serve as the direct object for the verb. This is critical because the logic of the experiment requires that there be more possible direct objects in the control conditions

than in the constrained conditions. To evaluate the number of verb-compatible objects in each scene, we asked six participants to identify how many items in each image could be direct objects of the constrained and control verbs paired with the image. Consistent with the design of the study, there were more possible direct objects in the control conditions ($M=6.48$) than in the constrained conditions ($M=2.44$) ($F(1,14)=62.21$; $p<.05$). Although there was no effect of information type, there was a significant interaction between information type and constraint. Simple effects analysis showed that participants provided more potential direct objects in the event-constrained condition ($M=3.38$) than in the verb-constrained condition ($M=1.51$) ($p<.05$). This difference likely occurred because by design, the verbs in the event-constrained condition were only weakly constraining, so it was more likely that more items in the scene could serve as their direct objects. This could be a concern if the presence of more potential direct objects led to more competition and less looking to the target; this could cause target looks in the event-constrained condition to be slowed relative to the verb-constrained condition. We will return to this issue in the Discussion. Although naturalistic scenes introduced variability to our stimuli, their enhanced situational information and strong support of event knowledge made them the appropriate choice of visual stimulus for the current experiment.

Procedure

36 undergraduate students from the University of Pittsburgh who had not participated in the norming completed the experiment for course credit. Participants' eyes were tracked using an Eyelink 1000 tracker (SR Research Ltd., Toronto, Ontario, Canada) with a sampling rate of 1 ms. Participants viewed stimuli binocularly on a monitor approximately 63 cm from their eyes. Head movements were minimized using forehead and chin rests. The experiment began with instructions and a 13-point calibration. A single-point centrally-located drift correction was performed after every trial, as well as a full 13-point recalibration every 24 trials; this drift correction ensured that participants were fixating the center of the screen when the visual stimulus was initially displayed. On each trial, the visual stimulus preceded the audio stimulus by 1000 ms, giving participants time to extract event-related information from the scene. Audio stimuli were presented to participants via two speakers positioned at either side of the viewing monitor. The experiment lasted between 20 and 30 minutes.

We constructed two lists of stimuli for counterbalancing purposes. Each participant saw every visual stimulus, but heard only one of the two possible accompanying audio stimuli. Stimuli were presented to participants in random order.

Results

As previously discussed, analyses were over 30 items. Because the nature of our visual stimuli (naturalistic scenes instead of clip-art images) does not allow inclusion of explicit distractor objects, our critical comparisons were between conditions rather than between target and distractor objects in the same scene (Staub et al., 2012). We removed trials containing fixations longer than 2000 ms (12 trials), as well as trials in which the target was never fixated (93 trials). This resulted in 9.7% of trials being discarded; 975 total trials were analyzed.

Linguistically-Binned Analyses

We begin with an analysis of proportions of fixations to the target in four different time bins: a bin lasting the duration of the verb and determiner, a bin lasting the duration of the noun, and two 500-ms bins beginning after noun offset; these last bins were roughly equal in duration to the two initial bins, and were intended to capture any changes in target fixation patterns after the end of the audio stimulus. Bins were generated for each item individually, so that different durations of verbs, determiners and nouns across items were accounted for. Fixations were considered to belong to a time bin if they started within that bin. This analysis therefore captured the way visual attention shifted to the target in reaction to the linguistic stimulus participants were hearing. Its time course is relatively coarse-grained, but the data in the verb (and determiner) bin provide a measure of all anticipatory gazes.

Fixations to the target were analyzed using empirical logit linear mixed-effects models grouped by participants, with information type (event vs. verb) and constraint (constrained vs. unconstrained) as fixed effects. Random intercepts of participant and item were also added to the models, as were random slopes of information type and constraint within participants. Data were grouped by participants because the object of these tests was to determine whether the results would generalize to other groups of participants. All results using linear mixed-effects models were also evident in ANOVAs. Figure 2 shows fixation proportions to the target for each condition in each bin.

Verb Bin—Following standard practice in the field (e.g. Mack, Ji, & Thompson, 2013), we assumed a 200ms lag between information uptake and saccades informed by that information, and so shifted our bins by 200ms. The verb bin therefore was from 200ms post verb onset to 200ms post determiner offset. Critically, it is unlikely that offsetting the bins in this way caused important pre-200 ms effects to be overlooked: a finer-grained time-course analysis to be presented below indicated that no differences between constrained and control conditions emerged until well into the verb. Eye movements to the target during this bin are a measure of predictive processing based on the constraints of the verb in the linguistic stimulus and qualities of the scene depicted in the image. Participants were significantly more likely to look at the target in the two constrained conditions than in the two control conditions ($\beta=-0.299$; $SE=0.12$; $t=-2.487$; $p<.05$), suggesting that they were able to use both types of information to make predictions. There was also a significant effect of information type in this bin: participants looked at the target more often in the verb conditions than the event conditions ($\beta=-0.643$; $SE=0.072$; $t=-6.511$; $p<.05$). This difference could be related to the fact that different images were used for the verb and event items, or to the fact that in the sentence+picture norm, participants were slightly more likely to respond with the target in the verb conditions than in the event conditions. Critically, there was no interaction of information type and constraint ($\beta=-0.197$; $SE=0.197$; $t=-1.001$; $p=.32$), suggesting that at least in this coarse-grained bin consisting of the verb and determiner, constraining information increased the proportion of looks to the target similarly in both the verb and event conditions.

Noun Bin—The noun bin began 200 ms after the onset of the noun and ended 200 ms after the offset of the noun. There was no effect of constraint ($\beta=0.088$; $SE=0.091$; $t=0.972$; $p=.$

335). Participants were more likely to look at the target in the two verb conditions than the two event conditions ($\beta=-0.363$; $SE=0.089$; $t=-4.093$; $p<.05$), similar to the previous bin. Finally, there was no interaction of information type and constraint ($\beta=-0.103$; $SE=0.176$; $t=-0.589$; $p=.557$).

First Post-Noun Bin—The first post-noun bin began 200 ms after the offset of the noun and ended 700 ms after the offset of the noun. A model containing the random slope of information type within participants did not converge, and therefore only a random slope of constraint within participants was used. Participants were more likely to look at the target in the two control conditions than in the constrained conditions in this bin ($\beta=0.352$; $SE=0.125$; $t=2.823$; $p<.05$). There was no effect of information type ($\beta=-0.086$; $SE=0.121$; $t=-0.711$; $p=0.479$) and no interaction ($\beta=0.116$; $SE=0.242$; $t=0.479$; $p=0.633$).

Second Post-Noun Bin—The second post-noun bin began 700 ms after the offset of the noun and ended 1200 ms after the offset of the noun. There were no effects of constraint ($\beta=0.070$; $SE=0.148$; $t=0.479$; $p=0.635$) or information type ($\beta=-0.099$; $SE=0.1278$; $t=-0.778$; $p=0.439$), and no interaction ($\beta=0.059$; $SE=0.254$; $t=0.236$; $p=0.814$).

Finer-Grained Time-Course Analysis

The analysis of the verb time bin above shows that participants were using the information in both constrained conditions to anticipate the target even before it was encountered. However, this analysis does not provide detailed time course information about when those anticipatory gazes occurred, leaving it unclear whether anticipatory looks may have been launched earlier on the basis of one information type than another. In order to determine when looks to the target in the two constrained conditions diverged from looks to the target in their corresponding control conditions, we conducted a time-course analysis of proportions of target fixations separately for each information type over a window beginning 200 ms after the onset of the verb and ending 3000 ms later. This time window was divided into smaller bins of 200 ms each (models using smaller bins did not converge, but graphs using smaller bins showed the same patterns on visual inspection). The current analyses were time-locked to verb onset, so that early effects of verb information could be captured.

To justify analyses testing for effects of constraint within individual post hoc time bins, we began by looking for an interaction between constraint and time bin across the dataset. We used linear mixed-effects models with fixed effects of constraint (constrained vs. control) and time bin, and random intercepts of participant and item. Although we expected that proportions of fixations to the target would increase over time, we did not expect that this increase would be strictly linear. Therefore, following Mirman, Dixon and Magnuson (2008), we added a quadratic time bin component to the model. If significant, this component would indicate that proportions of fixations to the target change nonlinearly across time bins. Based on significant interactions between the quadratic time component and constraint (event conditions: $\beta=0.396$; $SE=0.158$; $t=2.502$; $p<.05$; verb conditions: $\beta=0.396$; $SE=0.158$; $t=2.502$; $p<.05$), it was justified to test the effect of constraint within time bins.

We tested for a significant effect of constraint for each 200 ms bin beginning 200 ms after verb onset. Both participants and items were added as random intercepts. P-values for each test are shown in Table 6.

Graphs of fixation proportions in each bin for the verb information type can be seen in Figure 3a. Participants began looking towards the target more in the verb-constrained condition than in the verb-control condition in the 600-800 ms bin ($p=.01$). There was a similar marginal effect of condition in the 1000-1200 ms window ($p=.07$). This effect reversed in the 1600-1800, 2000-2200, and 2400-2600 ms windows. In these windows, participants looked more towards the target in the control condition than the constrained condition, mirroring the results found in the first post-noun bin in the linguistically-binned analysis above.

Figure 3b shows fixation proportions in each bin for the event information type conditions. Participants began looking towards the target more in the event-constrained condition than in the event-control condition in the 400-600 and 600-800 ms bins. Similar to the pattern in the verb conditions, this effect reversed in the 1800-2000, 2000-2200, and 2400-2600 ms bins.

When interpreting these analyses, it is important to consider that norming indicated there was a marginally significant 40 ms effect of information type on verb duration that may have influenced when the critical information was available. To address this concern, we removed two items with unusually long verb durations from the verb conditions and verified that across this subset of the data, verb duration did not differ by information type or constraint ($ps \geq .3$). We re-ran the time-course analysis with these items removed. This analysis yielded the same pattern of results as did the analysis using the full dataset: looks to the target in the verb-constrained condition began to diverge from looks to the target in the verb-control condition in the 600-800 ms bin ($p < .01$), 200 ms later than in the event conditions. These findings suggest that the effects of information type were not driven by differences in verb duration across conditions.

Latency of the first fixation to target after verb onset time-locked to noun onset

For completeness we report one final analysis, namely the latency of the first fixation to the target after verb onset for each condition. For trials on which the participant was fixating the target during verb onset, we used the second fixation to the target (following Staub et al, 2012), as this was the first fixation that could be driven by verb information. The latencies of these first fixations were time-locked to noun onset by subtracting each fixation-onset time from noun-onset time. This means that anticipatory fixations that occurred before noun onset had negative latencies, whereas fixations that occurred after noun onset had positive latencies.

There are a number of considerations to take into account when interpreting this analysis. Although some previous visual world studies do find effects of verb constraint in latency measures (e.g., Altmann & Kamide, 1999), it is much more common for studies to find effects of verb constraints in fixation-proportion measures like those reported above (e.g., Kamide, et al., 2003; Borovsky, et al., 2012). Furthermore, although it might seem that latency would be the most accurate measure of *when* visual attention shifts, Staub et al

(2012) show that latency effects sometimes reflect differences in *where* visual attention is directed: the presence of verb constraints can reduce the number of fixations on non-target items, which can create the appearance of faster convergence on the target.

Figure 4 shows mean latency to fixate the target time-locked to noun onset. Fixations were analyzed using linear mixed-effects models with fixed effects of constraint (constrained vs. control) and information type (event vs. verb), and random effects of participant and item. We also included random slopes of information type for participants and of constraint for both participants and items. There was a marginal effect of constraint ($\beta=-90.99$; $SE=46.58$; $t=-1.95$; $p=.054$) such that latency to fixate the target was faster in the constrained conditions than the control conditions. This weak effect mirrors the effect of constraint that appeared robustly in the verb bin in the initial linguistically binned analysis of target fixation proportions. The effect of information type was not reliable ($\beta=-191.72$; $SE=114.33$; $t=-1.68$; $p=.10$), nor was there any hint of an interaction ($\beta=6.23$; $SE=107.56$; $t=0.06$; $p=.95$). Still there was a numerical trend such that latency to fixate the target was approximately 100 ms earlier in the verb-constrained condition than the event-constrained condition. Although this pattern provides a very weak hint that prediction might be occurring earlier in the verb-constrained condition than in the event-constrained condition, we are hesitant to put much weight on it because none of the results reach statistical significance, latency may not be as direct a measure of when visual attention shifts as it would seem, and these findings contradict the more robust effects that appear in the binned analysis.

Discussion

This study did not find any boost to direct object prediction associated with the presence of a constraining verb. Participants anticipated the direct object in both the event-constrained and verb-constrained conditions, as evidenced by the main effect of constraint on proportion of fixations to the target during the verb. In fact, a finer-grained time course analysis showed that looks to the target in the event-constrained condition began diverging from looks in the event-control condition earlier than the verb-constrained condition diverged from the verb-control condition. This suggests that participants may have predicted direct objects more quickly when predictions were driven by strong event-based information, rather than the combination of event-based information and strong verb constraints. It is worth noting that the faster emergence of predictive target looks in the event-constrained condition cannot be because there was more constraining information in this condition. The sentence+picture cloze norming indicated that participants were slightly more likely to provide the target in the verb-constrained condition, with both event-based and strong verb constraints. This finding of earlier divergence in the event conditions also suggests that the larger number of possible direct objects in the event-constrained condition than verb-constrained condition did not lead to increased competition and slower convergence on the target.

The fact that the presence of selectional restrictions did not result in more or faster prediction is strong evidence against theories in which selectional restrictions are used to guide language processing before world knowledge comes into play. The current results are consistent with previous evidence that event-based knowledge is used very early in language comprehension (McRae & Matsuki, 2009), but critically strengthen support for the argument

that event-based knowledge is used as quickly as selectional restrictions are in comprehension.

One might be concerned that the strong event knowledge constraints in both constrained conditions might have overwhelmed the selectional restriction constraints. Given that the intent of the experiment was to test whether selectional restrictions are used earlier in processing than world knowledge, it would be a problem if the event knowledge that drove prediction were available to participants considerably before the selectional restriction information, because participants might then have already made predictions and show no boost or speed up associated with late-arriving selectional restrictions. We think this is unlikely to be an issue in the current experiment for a number of reasons. First, cloze norming testing just the picture and “Someone will” showed that prior to the presentation of the verb, the stimuli used in the event and verb conditions were similarly unconstraining of the target. This was the case because the images used as stimuli critically did not instantiate the target events, but only provided a scenario in which the target events might plausibly take place. In the sentence+picture norm, which added the verb to the above norm, rates of target completion shot up in the constrained conditions, indicating that the verb was crucial to identifying the target event in both constrained conditions. These considerations suggest that the predictive constraints provided by both selectional restrictions and event knowledge were generated at the verb, putting them on a similar footing with respect to timing. It is also worth noting that in this norm participants generated the target most often in the verb-constrained condition, suggesting that strong verb constraints helped drive direct object prediction even in the presence of the non-linguistic semantic information provided by the image.

The findings of this experiment raise a number of questions about selectional restrictions. First, what are they? Given the current findings, and compelling logical arguments about the impossibility of partitioning lexical and non-lexical meaning (Clark, 1983; Jackendoff, 2002), one likely possibility is that selectional restrictions are a specialized form of world knowledge (Matsuki et al., 2011; Warren et al., 2015). However, they must be something more than knowledge of an individual event, because violating selectional restrictions leads to different patterns of disruption than violating simple event knowledge (Warren & McConnell, 2007; Warren et al., 2015). One possibility, suggested by Warren et al. (2015), is that selectional restrictions are verb-related abstractions across world knowledge (see Resnik (1996) for a computational model that implements a very similar idea). To account for patterns of ERP effects, Paczynski and Kuperberg (2012) and Kuperberg (2013) proposed that comprehenders generate and use event abstractions at both coarse-grained levels, like Agent <animate>; Action; Patient <inanimate>, and fine-grained levels, like Experiencer <people>; State <awe>; Stimulus <view>; Place <peak of mountain>. Selectional restrictions could be event abstractions linked to particular verb senses, which are specific with respect to the verb and coarser-grained with respect to its arguments. For example, Agent <animate>; Action <entertain>; Patient <sentient>. This account has the advantage of explaining the source of selectional restrictions, as well as allowing flexibility in exactly what semantic properties can contribute to selectional restrictions.

A second important question is: how can we resolve the apparent contradiction between the findings of this study and previous studies in which selectional restriction violations caused earlier disruption than world knowledge violations? One possibility, consistent with the arguments made by Paczynski and Kuperberg (2012), is that those timing and amplitude differences are functions of error signals, rather than indicators of differences in the availability of different kinds of knowledge. Perhaps violating coarse-grained abstractions generates a stronger, earlier-detected error signal than violating fine-grained world knowledge regarding specific events. In the current study, there were no violations or error signals, only facilitation. If fine-grained event knowledge plays a strong early role in facilitation and prediction (e.g. Kamide et al., 2003; Matsuki et al., 2011; see also Bicknell, et al., 2010), then additional facilitation based on coarser-grained selectional restrictions may simply be redundant. This is because every likely argument will necessarily be compatible with both fine-grained world knowledge and coarse-grained abstractions, and the fine-grained world knowledge is more constraining.

In sum, the processing consequences of selectional restrictions may tell us less about divisions between different types of meaning than about the kinds of abstractions we generate over experience. Further work will explore how those different kinds and levels of abstraction influence facilitative and integrative processes during comprehension.

Acknowledgments

This research was supported by the National Institutes of Health under grant R01DC011520 to the second and third authors.

Works Cited

- Altmann GT, Kamide Y. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*. 1999; 73:247–264. [PubMed: 10585516]
- Baayen, RH.; Piepenbrock, R.; Gulikers, L. Linguistic Data Consortium. Philadelphia: University of Pennsylvania; 1995. The CELEX lexical database.
- Borovsky A, Elman JL, Fernald A. Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *J Exp Child Psychol*. 2012; 112(4):417–436.10.1016/j.jecp.2012.01.005 [PubMed: 22632758]
- Chambers CG, Tanenhaus MK, Magnuson JS. Actions and affordances in syntactic ambiguity resolution. *J Exp Psychol Learn Mem Cogn*. 2004; 30(3):687–696.10.1037/0278-7393.30.3.687 [PubMed: 15099136]
- Chomsky N. Aspects of the theory of syntax. 1965 MASSACHUSETTS INST OF TECH CAMBRIDGE RESEARCH LAB OF ELECTRONICS.
- Clark, HH. Making sense of nonce sense. In: Flores d'Arcais, GB.; Jarvella, RJ., editors. *The Process of Language Understanding*. John Wiley & Sons Ltd; 1983.
- Ferretti TR, Kutas M, McRae K. Verb aspect and the activation of event knowledge. *J Exp Psychol Learn Mem Cogn*. 2007; 33(1):182–196.10.1037/0278-7393.33.1.182 [PubMed: 17201561]
- Ferretti TR, McRae K, Hatherell A. Integrating Verbs, Situation Schemas, and Thematic Role Concepts. *J Mem Lang*. 2001; 44(4):516–547.10.1006/jmla.2000.2728
- Hagoort P, Hald L, Bastiaansen M, Petersson KM. Integration of word meaning and world knowledge in language comprehension. *Science*. 2004; 304:438–441. [PubMed: 15031438]
- Jackendoff, R. *Foundations of language: Brain, meaning, grammar, evolution*. 1st. New York, NY: Oxford University Press; 2002. p. 275-293.

- Kamide Y, Altmann GTM, Haywood SL. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *J Mem Lang*. 2003; 49(1):133–156.10.1016/s0749-596x(03)00023-8
- Katz JJ, Fodor JA. The structure of a semantic theory. *Language*. 1963; 39(2):170–210.
- Kukona A, Fang SY, Aicher KA, Chen H, Magnuson JS. The time course of anticipatory constraint integration. *Cognition*. 2011; 119(1):23–42.10.1016/j.cognition.2010.12.002 [PubMed: 21237450]
- Kuperberg, GR. The Proactive Comprehender: What Event-Related Potentials Tell Us About the Dynamics of Reading Comprehension. In: Miller, B.; Cutting, L.; McCardle, P., editors. *Unraveling the Behavioral, Neurobiological, and Genetic Components of Reading Comprehension*. Baltimore: Paul: Brookes Publishing; 2013.
- Marslen-Wilson W, Brown CM, Tyler LK. Lexical representations in spoken language comprehension. *Lang Cogn Process*. 1988; 3(1):1–16.
- Matsuki K, Chow T, Hare M, Elman JL, Scheepers C, McRae K. Event-based plausibility immediately influences on-line language comprehension. *J Exp Psychol Learn Mem Cogn*. 2011; 37(4):913–934.10.1037/a0022964 [PubMed: 21517222]
- McRae K, Hare M, Elman JL, Ferretti TR. A basis for generating expectancies for verbs from nouns. *Mem Cognit*. 2005; 33(7):1174–1184.
- McRae K, Matsuki K. People use their knowledge of common events to understand language, and do so as quickly as possible. *Language and Linguistics Compass*. 2009; 3(6):1417–1429. [PubMed: 22125574]
- Paczynski M, Kuperberg GR. Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *J Mem Lang*. 2012; 67(4):426–448.10.1016/j.jml.2012.07.003 [PubMed: 23284226]
- Pylkkänen, Liina; Bridget, Oliveri; Smart Andrew, J. Semantics vs world knowledge in prefrontal cortex. *Lang Cogn Process*. 2009; 24(9):1313–1334.10.1080/01690960903120176
- Staub, Adrian; Matthew, Abbott; Bogartz Richard, S. Linguistically guided anticipatory eye movements in scene viewing. *Visual Cognition*. 2012; 20(8):922–946.10.1080/13506285.2012.715599
- Warren T, McConnell K. Investigating effects of selectional restriction violations and plausibility violation severity on eye movements in reading. *Psychonomic Bulletin & Review*. 2007; 14(4):770–775. [PubMed: 17972747]
- Warren T, McConnell K, Rayner K. Effects of context on eye movements when reading about possible and impossible events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2008; 34(4):1001–1010.10.1037/0278-7393.34.4.1001
- Warren T, Milburn E, Patson ND, Dickey MW. Comprehending the impossible: what role do selectional restriction violations play? *Language, Cognition and Neuroscience*. 2015:1–8.10.1080/23273798.2015.1047458

Figure 1.

a: Image Accompanying Event-Constrained and Event-Control Sentences

Event-Constrained sentence: “Someone will lick the envelope.”

Event-Control sentence: “Someone will need the envelope.”

b: Image Accompanying Verb-Constrained and Verb-Control Sentences

Verb-Constrained Sentence: “Someone will pop the balloon.”

Verb-Control Sentence: “Someone will enjoy the balloon.”

Figure 2. Mean fixation proportions for each condition during each time window

Figure 3.

a: Time-course plot of fixation proportions in the verb conditions. The zero point on the X-axis indicates the beginning of the Verb Bin, or 200 ms post verb onset. Each data point represents proportion of target fixations in each 200 ms bin.

b: Time-course plot of fixation proportions in the event conditions. The zero point on the X-axis indicates the beginning of the Verb Bin, or 200 ms post verb onset. Each data point represents proportion of target fixations in each 200 ms bin.

Figure 4. Mean latency to fixate target, time-locked to noun onset

Table 1
Experimental conditions and associated constraints

Condition	Selectional Restriction	Event Knowledge
Event-Constrained	- Selectional Restriction	+ Event Knowledge
Event Control	- Selectional Restriction	- Event Knowledge
Verb-Constrained	+ Selectional Restriction	+ Event Knowledge
Verb Control	- Selectional Restriction	- Event Knowledge

Table 2
Sentence Cloze Norm: Idealized and Actual Constrained Response Frequencies and Standard Errors per Condition

Condition	Ideal Mean	Actual Mean	Standard Error
Event-Constrained	0	.31	.03
Event Control	0	.19	.03
Verb-Constrained	1	.65	.03
Verb Control	0	.20	.03

Table 3
Sentence & Picture Norm: Idealized and Actual Mean Target Response Frequencies and Standard Errors per Condition

Condition	Ideal Mean	Actual Mean	Standard Error
Event-Constrained	1	.88	.03
Event Control	0	.29	.04
Verb-Constrained	1	.97	.01
Verb Control	0	.36	.04

Table 4
Event (Picture Only) Norm: Mean Target Response Frequencies and Standard Errors per Condition

Condition	Mean	Standard Error
Event	.34	.03
Verb	.30	.03

Table 5

Mean Verb Segment and Determiner Lengths

Condition	Verb Segment	Standard Error	Determiner	Standard Error
Event-Constrained	577	26.18	250	10.82
Event Control	571	20.04	248	11.39
Verb-Constrained	611	27.63	242	12.08
Verb Control	619	22.74	277	13.12

Table 6

Time-Course Analysis: P-values and Z-values For Constraint Effect In Each Bin After Verb Onset

Information Type	200	400	600	800	1000	1200	1400	1600	1800	2000	2200	2400	2600	2800	3000
Verb	p-value ###	.99	*.01	.13	.07	.91	.67	*.02	.49	*.01	.15	*.04	.39	.26	.85
	z-value ###	.02	2.57	1.52	1.79	-.12	.43	-2.35	-.68	-2.82	-1.44	-2.09	-.86	-1.14	.19
Event	p-value ###	.78	*.01	.49	.96	.31	.27	.33	*.03	*.01	.28	*.03	.35	.75	.68
	z-value ###	-.28	2.60	2.06	.69	1.02	-1.09	-.98	-2.18	-2.52	-1.08	-2.12	.93	-.32	.42

* p<.05

values could not be calculated