Journal of Memory and Language 69 (2013) 196–227

Contents lists available at SciVerse ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

What happened (and what did not): Discourse constraints on encoding of plausible alternatives

Scott H. Fraundorf*, Aaron S. Benjamin, Duane G. Watson

University of Illinois at Urbana–Champaign, Department of Psychology, Champaign, IL 61820, United States

ARTICLE INFO

Article history:
Received 3 October 2012
revision received 6 June 2013
Available online 17 July 2013

Keywords:
Discourse
Recognition memory
Reading
Alternative sets
Fonts

ABSTRACT

Three experiments investigated how font emphasis influences reading and remembering discourse. Although past work suggests that contrastive pitch contours benefit memory by promoting encoding of salient alternatives, it is unclear both whether this effect generalizes to other forms of linguistic prominence and how the set of alternatives is constrained. Participants read discourses in which some true propositions had salient alternatives (e.g., *British scientists found the endangered monkey* when the discourse also mentioned *French scientists*) and completed a recognition memory test. In Experiments 1 and 2, font emphasis in the initial presentation increased participants' ability to later reject false statements about salient alternatives but not about unmentioned items (e.g., *Portuguese scientists*). In Experiment 3, font emphasis helped reject false statements about plausible alternatives, but not about less plausible alternatives that were nevertheless established in the discourse. These results suggest readers encode a narrow set of only those alternatives plausible in the particular discourse. They also indicate that multiple manipulations of linguistic prominence, not just prosody, can lead to consideration of alternatives.

Introduction

Understanding and remembering a discourse may involve representing not only what *is* true but also salient alternative propositions that are *not* true. For instance, Fraundorf, Watson, and Benjamin (2010) reported evidence that certain prosodic contours in spoken discourse, described in detail below, led listeners to encode information about a salient alternative in the discourse, such as *The Scottish knight won* as an alternative to *The English knight won* in (1b). This information helped listeners remember the events of the discourse. In particular, remembering that the Scottish knight lost the tournament helped listeners later reject *Scottish* as the winner. But, it did not affect their ability to reject an unmentioned item like *Welsh*, which was never part of a salient alternative in the discourse.

(1a) The English and the Scottish knights held a jousting tournament.
(1b) The ENGLISH knight won.

But while there is general evidence that important alternatives may be encoded as part of a discourse representation, it is unclear exactly how that set of alternatives is defined. Comprehenders might consider a relatively broad set of alternatives, such as alternative propositions related to any discourse entities in the same semantic category. Alternately, they might consider only those alternative propositions that are made particularly salient or plausible in the discourse.

It is similarly unclear what cues lead comprehenders to encode these alternative sets. Fraundorf et al. (2010) manipulated a feature that might be particularly apt to

* Corresponding author.
E-mail address: scottfraundorf@gmail.com (S.H. Fraundorf).

0749-596X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jml.2013.06.003

influence whether comprehenders consider a salient alternative: the type of prosodic pitch accent. The ToBI system for prosodic transcription (Beckman & Elam, 1997; Silverman et al., 1992) distinguishes multiple types of prosodic pitch accents that are argued to have different meanings. For instance, H* pitch accents consist of a single high pitch target (H) on the stressed syllable of the word (*) and are broadly associated with new information. L + H* pitch accents consist of a low pitch target (L) followed by a rise to a high pitch target on the stressed syllable, are acoustically more prominent (Selkirk, 2002), and have been argued to have a contrastive reading (Pierrehumbert & Hirschberg, 1990; see also Gussenhoven, 1983, for similar arguments in a somewhat different system of prosodic transcription).

Representation of salient alternatives in memory might be expected to occur only when listeners hear particular prosodic cues such as the L + H* accent—and not, for instance, in written text. However, readers may generate implicit prosody even when reading silently (Breen & Clifton, 2010; Fodor, 1998). Moreover, other theoretical accounts (Calhoun, 2009) have proposed that when linguistic material is made more prominent than expected in *any* way, it brings to mind relevant alternatives. Under both these accounts, the representation of salient alternatives in memory could be observed even in written discourse as a function of cues other than overt prosody.

In three experiments, we investigated whether salient alternative propositions are represented in memory when reading written discourse, and tested how the set of alternatives is constrained by readers. Experiments 1 and 2 tested whether the specific memory benefits previously observed for L + H* pitch accents could also be obtained in written discourse as a function of a different manipulation: italicization or capitalization, which we refer to collectively as *font emphasis*.[1] Experiment 3 tested whether the alternatives represented included *any* possible alternative proposition that could be formed from the discourse or only those that, given the scenario in the discourse, were particularly likely alternatives. Experiments 2 and 3 also collected measures of reading time to determine whether the benefits of font emphasis were contingent on how readers initially read the emphasized words.

Representing alternatives in a discourse

Success in discourse comprehension may involve representing not only particular referents and propositions, but also calculating and representing a set of one or more salient alternatives (e.g., Blok & Eberle, 1999; Calhoun, 2009; Pierrehumbert & Hirschberg, 1990 Rooth, 1992; Selkirk, 1995; Steedman, 2000). For instance, Rooth (1992) has argued that placing a linguistic constituent in focus introduces a *focus semantic value,* which expands the semantic interpretation of a sentence by introducing a set of one or more alternative propositions.[2] The alternative set has been formalized as a set of propositions that could have been formed had the focused element been replaced by something else, but that, given the actual proposition, cannot be true (e.g., Rooth, 1992; Steedman, 2000; Van Deemter, 1999).

For example, (2) expresses the proposition that elaborative rehearsal should be used to maximize cognition. (Capitalization indicates an L + H* pitch accent.) According to accounts of contrastive prosody (e.g., Pierrehumbert & Hirschberg, 1990), the contrastive L + H* accent on *elaborative* conveys that it is indeed *elaborative* rehearsal rather than some other form of rehearsal that should be used to maximize cognition. That is, it relates the true proposition to a set of alternative propositions, such as *If you want to maximize cognition, use maintenance rehearsal*, that could have been formed by replacing *elaborative* with other potential modifiers for *rehearsal*. Speakers and writers may introduce alternatives in this manner for multiple reasons, commonly including the desire to call attention to the contrast between two outcomes (Gundel, 1999; Rooth, 1992).

(2) If you want to maximize cognition, use ELABORATIVE rehearsal.

Some evidence suggests consideration of alternatives can benefit how a discourse is represented and remembered over the long term. Fraundorf et al. (2010) presented participants with spoken discourses in which the first part of the discourse, which we term the *context passage*, established two pairs of items. For example, (3a) below mentioned *British* and *French* as one pair and *Malaysia* and *Indonesia* as a second pair. (Fraundorf et al. included two pairs per discourse to test whether improving memory for one pair would come at a cost to memory for the other, but no such effects were observed.) A second part of the discourse, which we term the *continuation*, mentioned one member of each pair. For instance, (3b) mentions *British* from the first pair and *Malaysia* from the second pair.

(3a) Both the British and the French biologists had been searching Malaysia and Indonesia for the endangered monkeys.

(3b) Finally, the British spotted one of the monkeys in Malaysia and planted a radio tag on it.

Each pair was created so that a proposition in the continuation would have a salient alternative. For example, the pairing of *British* and *French* in (3a) makes it plausible that the French, rather than the British, could have spotted the monkey in (3b). In the accounts proposed by Rooth (1992), Steedman (2000), and others, the salient alternative *The French spotted one of the monkeys* should thus be considered if *British* is made prominent in (4b).

Fraundorf et al. (2010) tested whether listeners encoded the salient alternative by manipulating whether the key word in the continuation received the L + H* pitch

[1] The term *font* is commonly used to refer to differences between type *faces* such as Arial and Times New Roman, but more properly refers to individual variations within a particular face ("Font", 2011). We follow the latter usage by referring to manipulations such as italicization as *font* changes.

[2] Formally, the focus semantic value includes both the true proposition and the alternatives (Rooth, 1992).

accent described above, which is argued to contrast one item with an alternative or alternatives, or received the H* pitch accent, which is more broadly associated with new information (Pierrehumbert & Hirschberg, 1990). The contrastive L + H* pitch accent indeed appeared to lead participants in the Fraundorf et al. (2010) study to actively encode and remember a salient alternative to the truth. A day after listening to the stories, participants took a recognition memory test that included three types of probes: correct statements, such as (4a), *alternative probes* that referred to items that were part of the original pairing in the discourse, such as (4b), and *unmentioned probes* that referred to items never present in the original discourse, such as (4c).

(4a) The British scientists found the endangered monkey.
(4b) The French scientists found the endangered monkey.
(4c) The Portuguese scientists found the endangered monkey.

Both the alternative and unmentioned probes were false statements that were to be rejected, but they differed in whether or not they expressed a proposition that was likely to be in the alternative set. Alternative probes referred to the salient alternative in the original discourse and should be part of any set of alternatives to the true proposition, whereas the unmentioned probes referred to a brand new item and should not. Hearing the contrastive L + H* accent, rather than an H* accent, in the original continuation improved participants' ability to reject the alternative probes. Crucially, however, it did not improve rejections of the unmentioned probes. This suggests that the memory benefit of the contrastive L + H* accent over the H* accent was not simply due to improved encoding of the target item, which could allow it be more easily discriminated from *any* false probe. Rather, the difference between the L + H* and H* likely involved listeners representing something about the salient alternative, such as remembering that *it was not the French scientists who found the monkey*. Such a representation would allow the salient alternative to be more easily rejected, but knowing that the French scientists did not find the monkey would not help reject a proposition relating to an unmentioned item (*the Portuguese scientists found the monkey*).

How are alternative sets constrained?

Although there is general evidence that alternatives in a discourse are represented and remembered, one unresolved issue is *which* alternatives comprise the alternative set. In most cases, it seems unlikely that comprehenders would consider every possible alternative proposition. For example, (5b) is an alternative to (5a) below in that it is formed by replacing the constituent *British* and constitutes a proposition that cannot be true if (5a) is true. However, it seems implausible that comprehenders would actually consider and encode (5b) as an alternative in online language comprehension. (Indeed, the results of Fraundorf et al. 2010 support this claim, as the contrastive L + H* accent did not benefit rejections of entirely unmentioned items.) Thus, most theoretical accounts (e.g., Rooth, 1992; Steedman, 2000; Van Deemter, 1999) have assumed that the alternative set must be constrained by context.

(5a) The British scientists found the endangered monkey.
(5b) The Martian scientists found the endangered monkey.

How the context constrains the alternative set, however, is unclear. One hypothesis is that context only loosely restricts the alternative set. Comprehenders might consider as alternatives any proposition that could be created by replacing the prominent item with some other item of the same semantic or syntactic category. This possibility is consistent with models in which the selection of an alternative set is strongly influenced by hierarchical taxonomies of concepts (e.g., Blok & Eberle, 1999): any discourse element that is part of the same superordinate category may be a relevant alternative. It is also consistent with effects in other linguistic domains that have been attributed to givenness or presence in the discourse (e.g., Bock & Mazzella, 1983; Halliday, 1967; Haviland & Clark, 1974).

A different hypothesis is that the set of alternatives is constrained not just by generic knowledge but by a *situation model* of the particular state of affairs described by an individual discourse (Zwaan & Radvansky, 1998). Some experiments suggest that the situation to which a discourse refers can narrowly restrict sets of relevant alternatives. For instance, referring expressions that are in principle ambiguous (e.g., *the green block* when multiple green blocks are present) can be unambiguously interpreted as referring to a single referent if the alternatives are task-irrelevant or physically distant, preventing them from being considered (Brown-Schmidt & Tanenhaus, 2008). The discourse can provide such constraints even when it does not concern physically copresent objects: electrophysiological evidence suggests that if one of two characters in a story is described as leaving the room, an otherwise ambiguous referring expression is treated as unambiguous (Nieuwland, Otten, & Van Berkum, 2007). Although these experiments have focused on how the *correct* referent of a referring expression is identified, they also suggest that the set of alternatives considered is relatively small and is constrained by the particular scenario described by the discourse. Further evidence that the situation model, and not merely generic knowledge, can define alternative sets comes from evidence that context can introduce *new* members into an alternative set from outside the same taxonomic category (Byram Washburn, Kaiser, & Zubizarreta, 2011).

The original experiments by Fraundorf et al. (2010) could not tease apart these two hypotheses about how alternative sets are constrained. In their recognition memory test, correct rejections of a highly salient alternative were compared to correct rejections of a completely unmentioned referent. As noted above, the contrastive

L + H* accent benefited the former type of correct rejection but not the latter, suggesting its effect lay in encoding of the salient alternative. However, this benefit would have obtained regardless of whether the alternative set was fixed narrowly to include only the most plausible propositions given situation model or fixed more broadly to include propositions referring to *any* discourse element of the same semantic category. Both of these sets would have excluded the completely unmentioned item.

In other situations, how an alternative set is constrained would have consequences for how a discourse is remembered: Some propositions may refer to items that are *mentioned* in the discourse but that may be less likely alternatives given the situation model. For example, consider the context passage (6a) and the continuation (6b) below. Both *Saturn* and *Neptune* are mentioned in the context passage, but differ in their relationship to the true proposition regarding *Jupiter* in (6b). *Saturn* is mentioned in the context passage as part of the same pair as *Jupiter* and is likely part of a salient alternative proposition (i.e., the photos taken of Saturn could have been lost instead). However, the discourse establishes that *Neptune* is a less likely alternative for *Jupiter* in (6b) because the mission to Neptune had not yet occurred.

(6a) Originally, the space probe Cosmo III was designed to fly past Jupiter and Saturn and send photos and measurements back to NASA from both planets. NASA needed this information to guide the videos they were going to take of Neptune on a future mission.

(6b) However, due to a glitch in the programming of the Cosmo III, it lost the photos taken of Jupiter and put the future mission in trouble.

Discourses like this provide an avenue for testing what constrains the set of encoded alternatives. If the alternative set comprises only the alternative or alternatives that are particularly plausible given the situation model, then encoding an alternative set should benefit later rejections of a false proposition involving the plausible alternative (e.g., *Saturn* in the above example), but not one involving a mentioned but implausible alternative (*Neptune*). However, if the alternative set is defined to encompass all given referents of the same semantic category, then both types of correct rejections should benefit from encoding the alternative set.

Emphasis in written text

A second unresolved issue is what cues lead comprehenders to encode an alternative set, and whether these effects are unique to spoken prosody. Although the memory effects observed by Fraundorf et al. (2010) were obtained using contrastive L + H* pitch accents, a generalization to other manipulations, including those in written text, would be predicted by multiple accounts of prosody. One account of the memory effects observed by Fraundorf et al. (2010) is that the meaning of specific pitch contours such as the L + H* to introduce or call to mind salient alternatives (Gussenhoven, 1983; Pierrehumbert & Hirschberg, 1990). Although the L + H* accent is not explicitly presented in a written discourse, readers appear to generate implicit prosody when reading silently (Breen & Clifton, 2010; Fodor, 1998) and linguistic devices in written discourse could plausibly lead readers to implicitly generate a L + H* pitch contour. A contrastive effect in written text would also be predicted by an account (e.g., Calhoun, 2009) in which contrastive interpretations are not the inherent meaning of particular linguistic categories but occur probabilistically. The greater the linguistic *prominence* of a word relative to expectations, the more likely it is to bring to mind a set of salient alternatives (Calhoun, 2009), where prominence[3] is broadly defined to include prosodic, syntactic, discursive, and other variables that indicate the importance of linguistic elements to a reader or listener (Birch & Rayner, 2010). Such a theory also predicts that manipulations in written text that increase the perceived prominence or importance of particular elements of the text would lead to the representation of salient alternatives.

Thus, it is plausible that readers could encode and remember salient alternatives in reading text. However, many prior studies on the processing of salient alternatives in text have focused on online processing or on metalinguistic judgments, and it is unclear if the influence on long-term memory for a discourse requires the explicit presentation of contrastive L + H* pitch accents. In the present study, we investigated whether font emphasis would lead to the representation of salient alternatives in memory.

Some existing evidence does suggest that font emphasis can indicate contrast. McAteer (1992) asked participants to freely describe the "meaning" of capitalized text and of italicized text and found that the most commonly used word in describing italicization (but not capitalization) was "contrast". To date, however, findings have been mixed as to whether this apparent contrastive interpretation actually benefits comprehension. Emphasizing text using font changes has sometimes been observed to improve memory for a variety of materials, including short discourses (Sanford, Sanford, Molle, & Emmott, 2006), confusable drug names (Filik, Purdy, Gale, & Gerrett, 2006) and science texts (Golding & Fowler, 1992; Lorch, Lorch, & Klusewitz, 1995). But in other cases, no benefit of font emphasis has been observed (Harp & Mayer, 1998). The hypothesis that font emphasis leads readers to encode an alternative set provides one explanation for these inconsistencies: remembering salient alternatives would benefit

[3] One way of characterizing prominence within an utterance has been in terms of the *focus*, the part of the sentence intended to contribute new information (Halliday, 1967). However, prominence might vary even among constituents that are not focused or among constituents within the focus (Birch & Rayner, 2010), and focus might actually constitute multiple orthogonal dimensions (e.g., Halliday, 1967; Steedman, 2000). In particular, some accounts have argued that contrastiveness may constitute a property separate from whether information is new and in focus (e.g., Gundel, 1999; Steedman, 2000). Thus, following Birch and Rayner (2010) and Calhoun (2009), we thus adopt the more general term *prominence* to refer to differences in importance that may not necessarily be isomorphic to focus.

performance on some memory tests—those that required ruling out those alternatives—but not others.

It should be noted that prior studies of font emphasis have also frequently differed in the specific font changes used, and particular font manipulations may differ in their effectiveness or in their interpretation. For example, Filik et al. (2006) found that capital letters benefited memory more than did colored text. McAteer (1992) found that participants frequently used the word "contrast" to describe the meaning of italicization but rarely used this word to describe capitalization, although this metalinguistic task does not necessarily reflect differences in online interpretation. To assess the generality of any effects in the present experiments across font manipulations, we separately tested two different manipulations: capital letters and italicization.

Reading time and depth of processing

An additional benefit of assessing comprehension using written discourses is that participants' reading time provides a measure of their initial, online processing of the discourse. The representation of alternative sets in memory may be contingent on the online processing of the emphasized material. For instance, Rooth (1992) has argued that fixing a focus semantic value, which includes the set of alternatives, is an optional process and not always performed. One plausible reason why readers might not always calculate an alternative set is that this, and other aspects of language processing, may be time-consuming. The need to interpret a sentence in time for the interpretation to be useful for the task at hand may prevent readers from spending the time to construct the most detailed linguistic representation possible (Ferreira & Patson, 2007; Sanford & Sturt, 2002). Indeed, research on human memory has established that, under time constraints, learners may not attempt to fully master all material (Son & Metcalfe, 2000; Thiede & Dunlosky, 1999). An important determinant of later memory is whether particular items preferentially receive additional study time (Dunlosky & Connor, 1997; Tullis & Benjamin, 2010). It is possible that the degree to which readers invest time in determining the alternative set when initially reading the discourse might partially explain variation in the accuracy of their later memory.

To date, however, findings are mixed as to whether the depth of discourse representations is indeed mediated by online processing. In some cases, longer reading times predict a greater probability of successful comprehension (Caplan, DeDe, Waters, Michaud, & Tripodis, 2011; Daneman, Lennertz, & Hannon, 2007). In other cases, no such relations are observed (Christianson & Luke, 2011; Reder & Kusbit, 1991; Ward & Sturt, 2007), which has led to the suggestion that deeper encoding does not necessarily require more online processing time (Ward & Sturt, 2007). Thus, it is unclear whether the representation of salient alternatives in memory would be predicted by readers' online processing of the original discourse.

Present work

In three experiments, we tested both whether font emphasis would lead readers to remember a set of one or more salient alternative propositions and how this set of alternatives is constrained.

In Experiment 1, we first tested whether font emphasis could improve rejections of a plausible alternative proposition about a discourse, as reported by Fraundorf et al. (2010) for contrastive (L + H*) pitch accents. To preview, both italics and capitals benefited memory in the same way that the L + H* accents did in the Fraundorf et al. study: they improved correct rejections of the salient alternative but not of a proposition about an unmentioned item. This pattern suggests that readers had encoded that particular alternative. Experiment 2 replicated this effect and also assessed whether or not representing the salient alternative in memory required additional online reading time.

Finally, in Experiment 3, we tested whether the set of alternatives is constrained only by prior mention in the discourse or also by plausibility in the situation model. We compared the benefits of font emphasis in rejecting two types of false statements: a false statement about an plausible alternative in the discourse and a false statement about referent that was mentioned in the discourse but that formed a less plausible alternative for the true proposition.

Experiment 1

We first tested whether font emphasis would benefit memory for discourse and whether those benefits would lie specifically in rejecting a salient alternative proposition, as reported by Fraundorf et al. (2010) for contrastive (L + H*) pitch accents. Recall that in those experiments, the key comparison was how a contrastive accent in discourses such as (3), reproduced below, affected later responses to three types of memory probes, reproduced below as (4).

(3a) Both the British and the French biologists had been searching Malaysia and Indonesia for the endangered monkeys.
(3b) Finally, the BRITISH spotted one of the monkeys in Malaysia and planted a radio tag on it.

(4a) The British scientists found the endangered monkey.
(4b) The French scientists found the endangered monkey.
(4c) The Portuguese scientists found the endangered monkey.

The crucial distinction in this design is between probes that expressed a salient alternative proposition from the original discourse, such as *French* in (4b), and probes that referred to an unmentioned item, such as *Portuguese* in (4c). Of course, it is likely that these two types of probes differ in their baseline attractiveness as lures: for instance, the unmentioned items are not seen during the study phase and are new to the discourse. The critical test of whether readers are encoding a salient alternative proposition, however, is whether correct rejections of each of the

two types of false probes differ in how they are affected by prominence in the original discourse. We detail this logic below.

One hypothesis is that the primary effect of the font emphasis is to enhance encoding of the correct proposition (i.e., *the British found the endangered monkey*). Enhancing memory for the correct proposition might also help readers reject the lures by process of elimination (the phenomenon of recollection rejection; Brainerd, Reyna, & Estrada, 2006; Matzen, Taylor, & Benjamin, 2011), but it should not do so exclusively for particular types of probes. That is, superior memory for the true proposition should help reject both the alternative lure (*French*) and the unmentioned lure (*Portuguese*).

However, on the basis of results from prosody (Fraundorf et al., 2010), we hypothesized that the effect of font emphasis is rather to promote encoding of a set of one or more salient alternative propositions that did not come true (i.e., encoding that *the French did not find the monkey*). Remembering *the French did not find the monkey* should help readers reject the false alternative probe (*French*). This benefit could potentially come about both through helping readers reject the alternative probe itself and through affirming the correct probe by process of elimination (because the alternative can be ruled out). Crucially, however, knowing that *the French did not find the monkey* does not help determine whether it was the British or the Portuguese who found the monkey. Thus, the hypothesis that font emphasis leads readers to encode an alternative set predicts that emphasis will benefit correct rejections of alternative probes (*French*) but not correct rejections of unmentioned probes (*Portuguese*).

In Experiment 1, we tested whether this pattern, previously observed with contrastive (L + H*) pitch accents in spoken discourse (Fraundorf et al., 2010), could also be observed as a function of two types of font emphasis in written discourse: capital letters and italics.

Method

Participants

Twenty-four individuals participated in partial fulfillment of a course requirement. In this and all other experiments, participants were native English speakers at the University of Illinois.

Materials

Participants read 36 discourses, taken from Experiment 3 of Fraundorf et al. (2010). Each discourse began with a context passage such as (3a), repeated below, that established two pairs of alternatives, such as the pair *British* versus *French* and the pair *Malaysia* versus *Indonesia*. A subsequent continuation passage mentioned one item from each pair of alternatives, such as *British* and *Indonesia* in (7). Across the set of discourses, an equal number of continuations referred to the member of the pair that the context passage had mentioned first (e.g., *British* was mentioned before *French* in the context passage) as referred to the member of the pair that the contrast passage had mentioned second.

During participants' initial reading of the discourse, some of the critical words in the continuation were displayed with font emphasis. Font emphasis was independently manipulated on each of the two critical words, so that, within a given passage, font emphasis could be on the first, the second, both, or neither of the critical words, resulting in the four conditions presented in (7) below. (We included the separate manipulation of two pairs per discourse for consistency with Fraundorf et al., 2010, but, as in their experiments, the properties of one pair of alternatives never had any effect on memory for the other pair of alternatives.)

(3a) Both the British and the French biologists had been searching Malaysia and Indonesia for the endangered monkeys.

(7a) Finally, the British spotted one of the monkeys in Malaysia and planted a radio tag on it.
(7b) Finally, the BRITISH spotted one of the monkeys in Malaysia and planted a radio tag on it.
(7c) Finally, the British spotted one of the monkeys in MALAYSIA and planted a radio tag on it.
(7d) Finally, the BRITISH spotted one of the monkeys in MALAYSIA and planted a radio tag on it.

The type of font emphasis used was manipulated between participants. For half of the participants, emphasized words were displayed in capital letters, and for the other half, emphasized words were in italics.

In the recognition memory test, each critical word was tested with a probe statement about what happened in the continuation passage. Three probes were constructed for each item by varying a single word in the probe statement. A *correct probe*, such as (4a) above, referred to the correct item and should be affirmed. An *alternative probe*, such as (4b), referred to the other member of the pair in the original discourse and should be rejected. An *unmentioned probe*, such as (4c), referred to an item from the same semantic category but that was never mentioned in the original discourse; these probes should also be rejected. Each participant saw only one of these three probes for each item.

Each probe referred to only one of the two pairs of alternatives from a particular discourse. For example, the probes in (4) query which scientists found the monkey but not where the monkey was found. This allowed each pair to be separately tested, resulting in a total of 72 test items. No font emphasis was ever used in the test probes.

The assignment of items to the probe type and to the emphasis conditions was counterbalanced across participants using a Latin Square design. This resulted in a $3 \times 2 \times 2$ design: probe type (correct, contrast, or unmentioned) $\times$ presence of emphasis $\times$ emphasis type (capitals or italics), with the first two variables manipulated within participants and the lattermost manipulated between participants. An advantage of this design is that each critical word always appeared in the same syntactic and discourse context, regardless of font emphasis or the probe type with which it would eventually be tested. This eliminates any possibility that the effects of font emphasis are due to con-

founds with syntactic position or with the content of the rest of the discourse.

Lists of the discourses and probe questions used in Experiment 1 are available in the Appendices of Fraundorf et al. (2010).

Procedure

The experiment was performed on a computer running MATLAB and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Participants were instructed that they would read some stories for a subsequent memory test. The format of the memory test was not described in advance.

Participants first completed a study phase in which each story was presented one at a time in a random order. Stories were displayed on a computer monitor in white Arial text against a black background. In Experiment 1, the entire discourse was displayed on the screen at the same time. The context passage and continuation passage were viewed as a single paragraph. Participants took as long as needed to read the discourse, and then pressed the space bar to advance to the next discourse. There was a 1000 ms delay between stories. When participants had read 18 of the 36 stories, a message informed them that they were halfway done and invited them to take a break before continuing.

After reading the last story, participants proceeded immediately to the test phase. In the test phase, probe statements appeared on the screen one at a time in a re-randomized order. Participants indicated whether they thought each statement was *true* or *false* by pressing one of two keys on the keyboard. Participants were told that they should reject a statement if they thought that any part of it was false.

Results

Analytic strategy

Memory performance has sometimes been assessed using the proportion of accurate responses, sometimes with the goal of comparing accuracy between true statements to be affirmed and false statements to be rejected. However, neither affirmations of true probes or rejections of false probes provide a pure measure of a single behavioral process, as both reflect a combination of (a) discrimination in memory between true and false information and (b) an overall preference to respond *true* or *false* elicited by particular participants, items, or conditions (Donaldson, 1996; Freeman, Heathcote, Chalmers, & Hockley, 2010; Wright, Horry, & Skagerberg, 2008). That is, accuracy might vary between true and false probes simply due to participants' overall preference to respond *true* or *false* or the tendency of particular discourses to elicit *true* or *false* statements, rather than participants' ability to discriminate one probe type from another. To separate participants' response bias from their actual sensitivity to which probes were true and which probes were false probes, we applied the theory of signal detection (Freeman et al., 2010; Green & Swets, 1966; Macmillan & Creelman, 2005; Wright et al., 2008), in which data are parameterized as the proportion of *true* responses. This framework allows a theoretical and empirical dissociation between *response bias* (the overall baseline rate of responding *true*) and *sensitivity* (an increased probability of responding *true* when the probe statement is actually true, also known as *discrimination*). Although our primary interest was in participants' sensitivity to whether particular probes were true or false, we report results for both sensitivity and response bias to present a complete description of the data. (To preview, font emphasis affected sensitivity and not response bias.)

We then analyzed the data using mixed effect logit models (Baayen, Davidson, & Bates, 2008; Jaeger, 2008; see also Wright et al. (2008), for applications to recognition memory). In these models, the log odds (or *logit*) of responding *true* are modeled on a trial-by-trial basis. We adopted these models rather than ANOVAs for three reasons. The primary motivation is that one of the goals of Experiments 2 and 3 was to investigate whether participants' online reading time predicted their later memory. Evaluating this hypothesis required an approach in which variation in reading time could be related to memory at the level of individual trials rather than an aggregation of all trials within a condition. Although it would be possible to divide reading time into a categorical variable for use in an ANOVA (e.g., with a median split), such dichotomization greatly reduces statistical power by discarding the variation within each category (Cohen, 1983). Mixed effect logit models provide an effective way to relate reading time to later memory because they model performance on individual items and can easily incorporate continuously varying predictors such as reading time. Although reading time was not measured in Experiment 1, we apply the same methodology to Experiment 1 for consistency and ease of comparison across experiments.

A secondary motivation for using mixed effects logit models is that in some conditions, the proportion of *true* responses was low. Treating such proportions as the dependent variable in an ANOVA would be inappropriate: proportions far from .5 are not normally distributed in that their mean and variance are related (Agresti, 2007, p. 9; Jaeger, 2008). By comparison, the logit has variance unrelated to its mean across the range of possible proportions (Jaeger, 2008).

Finally, as in many psycholinguistic studies, variability across both the sampled participants and sampled items was of interest (Clark, 1973), and mixed effect models allow both of these sources of variability to be incorporated into a single model (Baayen et al., 2008).

Model fitting and results

Mixed effect models can include both fixed effects, representing variables for which the particular levels are of interest, and random effects, variables with levels randomly sampled from a larger population. The random effects included the items (propositions being tested) and participants. The fixed effects tested in Experiment 1 were the probe type, presence of font emphasis in the original story, and emphasis type, as well as their interactions. The three probe types were analyzed using two planned comparisons implemented using a sum-coding system. The first planned comparison tested sensitivity in rejecting the unmentioned probes relative to the baseline (mean) rate of rejections; this comparison tests what confers a

benefit in rejecting items that were never part of the discourse. The second comparison tested sensitivity in rejecting the alternative probes relative to the baseline rate of rejections; this comparison tests what confers a benefit in rejecting salient contrast items. As our primary theoretical interest was in how emphasis would benefit participants' later ability to reject the different types of false statements, both of these planned comparisons were coded so that a positive value indicated more correct rejections (i.e., fewer erroneous *true* responses). The other predictors were simply mean-centered; doing so provides parameters corresponding to the main effects in an ANOVA analysis (i.e., collapsing across the levels of the other variables). All models were fit in the R Project for Statistical Computing using the *lmer* function of the *lme4* package (Bates, Maechler, & Bolker, 2011).

In each discourse, there were two propositions tested, each of which could be independently analyzed. It was possible that memory regarding one proposition (e.g., whether the *British* or *French* scientists found the monkey) could be influenced by whether the key word related to the other proposition (whether the monkey was found in *Malaysia* or *Indonesia*) was emphasized. However, a preliminary analysis indicated that this variable had no effect, consistent with past data from young adults on font emphasis (e.g., Lorch et al., 1995) and on pitch accents (Fraundorf et al., 2010; but see Fraundorf, Watson, and Benjamin (2012), for differing results in other populations, including older adults). Consequently, we collapsed across this variable for all subsequent analyses.

Proportion of *true* responses in each condition is displayed in Fig. 1 as a function of probe type, presence of font emphasis in the original discourse, and emphasis type.

In a mixed effect model, variability in an effect across participants or items can be represented with a random slope of that effect by participants or items. Random slopes by subjects for the two within-subjects factors (probe type and presence of font emphasis) did not improve the fit of the model, $\chi^2_{(20)} = 8.41$, $p = .99$. Random slopes by items for the effects of font emphasis and probe type did improve the model, $\chi^2_{(20)} = 209.45$, $p < .001$, but no random slopes for emphasis type (capitalization versus italicization) further improved the model (all $ps > .9$). Thus, we report results from the model with only random slopes by items for probe type, presence of emphasis, and their interaction, the maximal random effects structure justified by the data. (Estimates of the random effects from all three experiments are presented separately in Appendix E, as our primary theoretical interest was in the fixed effects.)

Fixed effect parameter estimates for the final model are displayed in Table 1. Overall, the odds of responding *true* rather than *false* were 0.59 (95% CI: [0.50,0.70]), which reliably differed from chance, Wald $z = -5.93$, $p < .001$. This tendency to respond *false* is appropriate given that there were more false probes than true probes. The first planned comparison indicated that the odds of rejecting an unmentioned probe were 41.25 times (95% CI: [23.04,73.78]) greater than participants' baseline tendency to response *false*, $z = 12.54$, $p < .001$, indicating that participants were sensitive to the fact that the statements about the unmentioned probes were false. Collapsing across the presence of font emphasis, the odds of rejecting an alternative probe were 1.88 times (95% CI: [0.96,3.69]) than participants' baseline rejection rate (the second planned comparison), $z = 1.85$, $p = .06$, indicating only marginal sensitivity in memory. (However, as will be seen below, such rejections were substantially enhanced when the critical word was emphasized.)

The crucial question was how responding would be affected by the font emphasis in the original discourse. Font emphasis did not induce an overall bias to respond *true*, $z = -.53$, $p = .60$. Rather, it facilitated rejection of the alternative probes; the odds ratio between rejections of an alternative probe and the overall rejection rate (contrast 2) were 3.70 times (95% CI: [1.51,9.08]) greater when the critical word was originally emphasized, $z = 2.85$, $p < .01$. However, font emphasis did not reliably benefit rejections of the unmentioned probes, $z = 1.17$, $p = .24$.

Finally, there was a marginal 3-way interaction of probe type, presence of font emphasis, and emphasis type, indicating that the benefit of emphasis in rejecting the alternative probes was stronger for capital letters than italicization, with the effect being 3.39 times (95% CI:

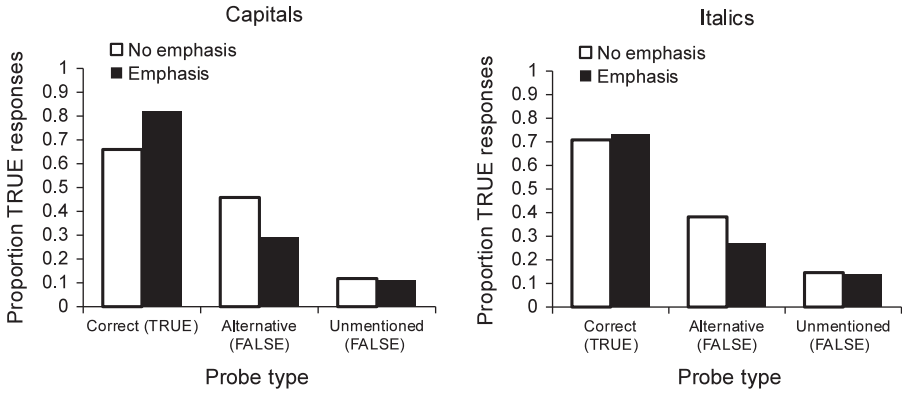

Fig. 1. Mean rate of *true* responses in Experiment 1 as a function of font emphasis and probe type, for participants who saw capitalization (top panel) and for participants who saw italicization (bottom panel). Responding *true* is a hit to a correct probe and a false alarm to an alternative or unmentioned probe.

Table 1
Fixed effect estimates for multi-level logit model of memory judgments in Experiment 1 (N = 1728, log-likelihood: −811).

Fixed effect	β	*SE*	Wald z	p
Baseline rate of *true* responses (response bias)	−0.53	0.09	−5.93	<.001
Emphasized word (effect on response bias)	−0.07	0.14	−0.44	.55
Rejections of unmentioned probe versus baseline (sensitivity)	3.72	0.30	12.54	<.001
Rejections of alternative probe versus baseline (sensitivity)	0.63	0.34	1.85	.06
Capitalization versus italics (effect on response bias)	0.08	0.14	0.60	.55
Emphasized word × capitalization (effect on response bias)	0.26	0.27	0.96	.34
Emphasized word × rejection of unmentioned probe (effect on sensitivity)	0.62	0.53	1.17	.24
Emphasized word × rejection of alternative probe (effect on sensitivity)	1.31	0.46	2.86	<.01
Capitalization × rejection of unmentioned probes (effect on sensitivity)	0.75	0.42	1.80	.07
Capitalization × rejection of alternatives (effect on sensitivity)	−0.44	0.36	−1.23	.22
Emphasized word × capitalization × rejection of unmentioned probes (effect on discrimination)	0.62	0.83	0.74	.45
Emphasized word × capitalization × rejection of alternative probes (effect on discrimination)	1.22	0.72	1.70	.09

Note. SE = standard error.

[0.83, 13.81]) greater for participants who saw capitals rather than italics, z = 1.70, p = .09.

Discussion

Experiment 1 demonstrated a mnemonic benefit of representing salient alternatives similar to that observed by Fraundorf et al. (2010). Placing font emphasis on one member of a pair of alternatives during the original presentation of a discourse improved participants' ability to reject an alternative proposition to the emphasized item on a subsequent memory test. However, font emphasis conferred no benefit in rejecting propositions about items that were never part of the discourse. This pattern suggests that the font emphasis led participants to encode a particular alternative to the true proposition, which would help reject that alternative but not items that were never in the alternative set. This result demonstrates that memory for a discourse can involve remembering not only what happened, but also important alternatives that did not happen.

Experiment 1 extended these prior results by demonstrating the effect is not limited to cases where participants hear contrastive (L + H*) pitch accents in spoken discourse. A similar benefit can also be observed in a different modality—written text—and with a different manipulation—font emphasis rather than pitch accents. This result is consistent with proposals that implicit prosody is generated in the process of silent reading (Breen & Clifton, 2010; Fodor, 1998), as well as with accounts in which any sufficiently prominent material brings to mind relevant alternatives (e.g., Calhoun, 2009).

Importantly, the pattern across conditions indicates that these mnemonic benefits cannot be attributed *only* to the perceptual properties of font emphasis. If memory for the discourse was improved simply because the emphasized words were perceptually salient or easier to read, this should have applied to any probe that tested memory for the information. However, the effect was more specific: emphasized text did not benefit rejections of a probe that referred to an item never in the discourse, suggesting the memory benefit lay in the encoding of an alternative set. Nevertheless, it is likely that the perceptual salience of emphasized text plays a role in its effects, and we return to this point in the Discussion.

The findings of Experiment 1 are qualified by the somewhat high rate of rejections of the unmentioned probes even in the absence of font emphasis (M = 13% affirmed, thus 87% rejected), which may have masked any potential benefit of font emphasis in rejecting them. In Experiment 3, we provide stronger evidence for representation of the salient alternatives by introducing a new type of false probe. The new false probe elicits a substantially higher rate of false alarms, but it still does not mention the alternative and still shows no effect of font emphasis.

Both capital letters and italicization promoted encoding of an alternative set in memory; in fact, the effect was stronger for capitalization. This conflicts with the finding that italicization is more apt to be described as having have a contrastive interpretation when participants are simply asked to describe the "meaning" of different kinds of font emphasis (McAteer, 1992). However, those metalinguistic judgments may not tap the same processes as reading and memory. Other evidence in fact suggests that metalinguistic judgments do not always predict the actual benefits of font emphasis: for instance, colored text is rated as more salient than capitalization, but capitalization is actually more effective at increasing memory for drug names (Filik et al., 2006). And, more generally, learners often incorrectly appraise which study conditions will lead to superior memory (for review, see Benjamin, 2005; Benjamin, 2008; Kornell & Bjork, 2007).

Experiment 1 thus provided evidence that in reading, as in spoken discourse comprehension, prominence of one element of the discourse leads comprehenders to encode a salient alternative proposition, formed by substituting the prominent element for a different element of the discourse, and that this representation benefits later memory. However, Experiment 1 did not reveal anything about how participants' initial, online processing of the discourse may have contributed to this memory benefit. In Experiment 2, we assessed participants' online reading time and how it did or did not relate to participants' later memory.

Experiment 2

Experiment 2 measured participants' online reading time while they read discourses containing two-item pairs similar to those in Experiment 1. Of particular interest was whether the memory benefit from the font emphasis was

contingent on how participants originally read the emphasized words. One hypothesis is that it may be time-consuming to calculate and encode the set of alternatives and that, as a result, readers do not always spend the time to do so (or do not do so to a degree that the alternatives can be remembered later). This hypothesis predicts that reading time on the emphasized word will be causally related to memory, with increased reading time predicting a greater likelihood of observing the memory benefit. By contrast, two other hypotheses do not predict a relationship between reading time and later memory. One such hypothesis is that fixing the set of alternatives requires no extra time. The other is that calculating the set of alternatives requires extra time, but that readers do so every time they encounter font emphasis. If readers *always* slow down to calculate the alternatives, reading time would not discriminate those trials on which the alternative set was encoded from those trials on which it was not. Thus, under either of these latter accounts, trial-by-trial variability in reading time would not predict later memory for the alternative. Experiment 2 pits these hypotheses against the hypothesis that readers only sometimes invest the time to calculate the alternative set.

Experiment 2 also provided a further test of the generality of the memory benefits by using different discourses than Experiment 1.

Method

Participants

Forty-eight individuals participated in partial fulfillment of a course requirement or for a cash honorarium. One of the original 48 participants did not complete the entire procedure within the 50 min allotted for the session and was replaced with an additional participant.

Materials

The Experiment 1 materials were substantially rewritten for Experiment 2 in order to add additional controls. First, readers are known to slow down at the ends of punctuation-marked sentences and clauses (for review, see Reichle, Warren, & McConnell, 2009), and this slowdown could overwhelm the effects of interest. To avoid this, the critical words in the continuation passage never appeared immediately before or after a punctuation mark between clauses.

Second, reading times increase at the start of a line and decrease at the end (for review, see Rayner, 1998). To ensure this effect did not vary across items, the discourses were written so that when the discourses were naturally spaced on the computer screen, the critical words never appeared first or last in a line.

Third, the two words in each pair (e.g., *Jupiter* and *Neptune*) were matched in number of characters. Readers are known to acquire information about the length of upcoming words before fixating them (for review, Rayner, 1998); matching the length of the two words in the pair prevented readers from obtaining information about the outcome of the discourse before actually reading the critical word itself. Because the unmentioned probe was not read in the original discourse, it was not necessary to control its length, but where possible, the unmentioned probe was also matched in character length to the other two probes as well.

Finally, in four discourses in Experiment 1, the referent that would be expected to be included in the alternative proposition was never mentioned explicitly and had to be pragmatically inferred. For example, in one discourse, *boys* was implicitly contrasted with *girls* without *girls* being mentioned in the context passage. Although some lexical items may inherently evoke relevant contrasts (Clifton, Bock, & Radó, 2000) that could become part of the alternative set (Pierrehumbert & Hirschberg, 1990; Rooth, 1992; Steedman, 2000), determining these alternatives may be more time-consuming for readers than when the alternative was explicitly introduced (Sedivy, 2002). Thus, variability in whether a salient alternative was explicitly introduced would likely introduce additional variability in reading time between items. In Experiment 2, the salient alternative item was always explicitly mentioned in the context passage, using the same lexical item that would appear in the continuation.

In all other respects, the items used in Experiment 2 had the same structure as those in Experiment 1. Probe type, presence of font emphasis, and emphasis type were manipulated using the same design as in Experiment 1.

A list of the Experiment 2 discourses appears in Appendix A and the test probes in Appendix B.

Procedure

In the study phase of Experiment 2, discourses were presented using the self-paced moving window paradigm (Just, Carpenter, & Woolley, 1982). The discourse was initially displayed on the screen with only the first word visible; the other words were replaced by lines. Participants pressed the space bar to advance through the discourses; after each press, the next word was displayed and the previous word replaced by a line. As in Experiment 1, the context passage and continuation passage were combined into a single paragraph.

In the moving window paradigm, text is most commonly presented in fixed-width faces such as Courier, in which every character occupies the same width on the screen. However, pilot testing suggested that participants found it difficult to detect italicization of the Courier face. We instead presented text in Arial, a face in which letters vary in their width. Words outside the moving window were replaced with lines exactly matched in length to the width on the screen that the words would occupy when presented in Arial.

To demonstrate that words in the experiment could be emphasized with font manipulations, one word in the initial instructions to participants was emphasized using the type of font emphasis (capitalization or italicization) that participants would later see in the experiment.

The recognition test procedure was unchanged between experiments.

Results

Due to an error in stimulus construction, for one memory probe the words used in the three probe conditions did

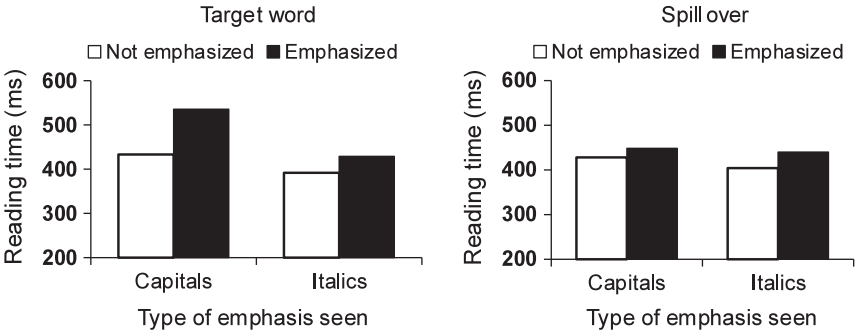

Fig. 2. Mean reading time in Experiment 2 on target words (left panel) and spillover words (right panel) as a function of font emphasis and emphasis type.

not match the referents named in the original discourse. We report results with data from this item omitted, but the inclusion or omission of this item did not affect any of the patterns described below.

Initial reading time

The characteristics of word *n* can also affect reading time to the following word *n* + 1 (Henderson & Ferreira, 1990; Rayner, 1998), so we examined reading times both on each critical word and on the *spillover* word that immediately followed. Mean reading times for each region and condition during the original presentation of the discourse are displayed in Fig. 2. Because the reading times were positively skewed (skewness = 7.09) and thus non-normal, we used the natural log of the reading times (skewness = 0.92) as the dependent variable in our models.

The model of reading time included random effects of participants and of items (words), as well as three fixed effects: region (critical word or spillover), presence of font emphasis (present or absent), and emphasis type (capitalization or italics), resulting in a $2 \times 2 \times 2$ design. Region and presence of font emphasis were coded using dummy coding. This coding system first tests the simple main effects of font emphasis within just the reference level (the critical word). Then, the interaction of other effects with the region variable tests whether those effects differed on the spillover word as compared to the critical word. Emphasis type (capitalization or italics) was mean-centered so that the main effects of region and font emphasis represent the mean of those effects across emphasis types. An interaction with emphasis type represents a stronger or weaker effect for one emphasis type relative to the other.

Random slopes for the two within-participants factors significantly improved the fit of the model, $\chi^2_{(9)} = 63.86$, $p < .001$, indicating some variability across participants in how much their reading times were affected by font emphasis. Random slopes by items for those same two factors further improved the fit of the model, $\chi^2_{(9)} = 32.24$, $p < .001$. The addition of random slopes by items of emphasis type (capitals or italics) and its interactions with the other factors did not improve the fit of the model any further, $\chi^2_{(26)} = 29.19$, $p = .30$.

Fixed effect parameter estimates from the final model are displayed in Table 2. First, consider the initial four parameters. These parameters test the simple main effects of font emphasis within the critical word itself.[4] The model revealed that, overall, emphasized words ($M = 462$ ms) were read more slowly than words without font emphasis ($M = 414$ ms), $t = 5.52$, $p < .001$. There was also an interaction with emphasis type: font emphasis increased reading times more for participants who saw capital letters rather than italics,[5] $t = 2.51$, $p < .05$.

The remaining parameters test whether the effects in the spillover region differ significantly from those in the critical region itself. There was a reliable three-way interaction of region type, font emphasis, and emphasis type, with the difference between capitalization and italicization disappearing by the spillover region, $t = -2.48$, $p < .05$.

Memory

One goal of Experiment 2 was to assess whether the benefit of font emphasis on discriminating the true proposition from the salient alternative proposition varied as a function of participants' online reading time. That is, was a combination of both font emphasis and increased reading time needed to obtain a mnemonic benefit? To test this hypothesis, we analyzed participants' memory performance using the same approach as in Experiment 1, but we added parameters for participants' initial reading time and its interactions with the other variables of interest.

One concern with using raw reading time as a predictor is that it confounds slowdown on the emphasized words with participant-level variation in reading speed. For instance, participants who were more motivated to remember the discourses may have both read more slowly and been more apt to calculate an alternative set. This could lead to an association of reading time with memory performance even if there were no causal relation between increased reading time and calculation of the focus semantic value.

[4] An alternate means of hypothesis testing in multi-level models with continuous dependent variables is Markov chain Monte Carlo simulations (Baayen et al., 2008). At present, however, these methods are not implemented for models containing random correlation parameters. Instead, Baayen (2008, p. 270) suggests using t values greater than 2 as a heuristic for significance at the $\alpha = .05$ level, since for models with thousands of observations, the t distribution has essentially converged to the normal distribution.

[5] Additional two-way analyses revealed that the effect of font emphasis on reading time was significant both within only those participants who saw capitals and only those participants who saw italics.

Table 2
Fixed effect estimates for multi-level model of log reading time in Experiment 2 (N = 6816, log-likelihood: −2710).

Fixed effect	β	SE	t	p
Non-emphasized critical word (baseline)	5.90	0.04	138.37	<.001
Emphasized word	0.10	0.02	5.52	<.001
Emphasis type is capitalization (versus italics)	0.06	0.08	0.66	.51
Emphasized word × emphasis type is capitalization	0.09	0.04	2.54	<.05
Spillover region	0.02	0.01	1.41	.16
Spillover region × emphasized word	−0.03	0.02	−1.48	.14
Spillover region × emphasis type is capitalization	>−0.01	0.02	−0.25	.80
Spillover region × emphasized word × capitalization	−0.10	0.04	−2.44	<.05

Note. SE = standard error.

A solution is to examine only within-subject differences in reading time. We calculated residual reading time (Ferreira & Clifton, 1986) by regressing, separately for each participant, reading time on (a) an intercept representing baseline reading speed and (b) the length[6] of each word. Residual reading time is the reading time left unexplained by these more basic factors. To obtain the most precise estimate of participants' reading speed, the regression models for calculating residual reading time included all words in the materials, not just the critical regions. Although residual reading time has typically been calculated from untransformed reading times, reading times, as noted above, are positively skewed, so we instead modeled log-transformed reading time.

We then analyzed the log odds of *true* responses as a function of probe type, presence of font emphasis, and emphasis type, as well as residual reading time. Residual reading time was summed over the critical and spillover words. One percent of the observations contained residual reading times more than three standard deviations from a participant's mean in that condition. These outlying observations might reflect extraneous influences irrelevant to the task, such as sound from outside the experiment or a participant sneezing. Mixed effects models are robust against missing data (Quené & van den Bergh, 2004), so rather than make potentially erroneous assumptions about what outlying reading times should be replaced with, we simply eliminated these observations, affecting 1% of the data. The predictor variables were again coded using mean centering to obtain estimates of the main effects. Reading time was also mean-centered; consequently, the main effects of other variables represent effects of those variables at an *average* residual reading time for the critical region.

Fixed effect parameter estimates from the model are displayed in Table 3. A preliminary analysis indicated that, once reading time during the study phase was accounted for, emphasis type (capitals or italics) made no further contribution to the model, $\chi^2_{(12)} = 10.30$, $p = .59$, so we dropped this variable to simplify the model. However, the model was improved by the inclusion of random slopes for probe type by participants, $\chi^2_{(5)} = 23.76$, $p < .001$, and by items, $\chi^2_{(5)} = 424.25$, $p < .001$. No other random slopes approached significance (all $ps > .25$).

Mean rates of *true* responses in each condition are displayed in Fig. 3. The memory effects observed in Experiment 1 were replicated. At a mean level of residual reading time, font emphasis improved readers' rejection of the alternative probes: the presence of font emphasis increased the odds ratio of rejections of the alternative probes over the baseline response bias (the second planned comparison) by 1.63 times (95% CI: [1.01, 2.64]), $z = 1.97$, $p < .05$. By comparison, font emphasis did not benefit rejections of the unmentioned probes (the first planned comparison), $z = -0.53$, $p = .60$. In fact, the effect was numerically in the opposite direction, with font emphasis decreasing the odds of this rejection by 0.87 times (95% CI: [0.49, 1.63]).

Additionally, Experiment 2 found that the benefit of font emphasis in rejecting the alternative probes was amplified by increased reading time. The model revealed a three-way interaction between font emphasis, probe type, and reading time. An increase of 1 log millisecond[7] of reading time on emphasized material corresponded to a 2.69 times increase in the odds ratio between rejections of alternative probes and the baseline rejection rate (95% CI: [1.04, 6.91]), $z = 2.05$, $p < .05$. That is, the more that readers slowed down on the emphasized words, the more likely font emphasis was to improve rejections of the alternative probes.

Again, this effect was limited to rejecting the salient alternative proposition, not all false statements. A 1-unit increase in residual reading time on emphasized material actually resulted in a numerical decrease in the odds of rejecting the unmentioned probe by 0.77 times (95% CI: [0.27, 2.20]), although this effect was not statistically reliable, $z = -0.49$, $p = .62$. The two-way interaction between reading time and rejecting the alternative probe, collapsing across the presence of emphasis, was also not significant, $z = -0.48$, $p = .63$. That is, increased reading time was not generally predictive of improved rejection of alternative probes. (In fact, the relation was numerically

[6] This method assumes a linear relationship between word length and the dependent variable. To test if this assumption was valid, we first fit an additional mixed effects model. This model included all words in the experiment and modeled log reading time as a function of both linear and quadratic effects of word length, as well of a random slope for word length by participants and a random intercept for both participants and items. The linear effect of reading time was a robust predictor of reading time, $t = 9.45$, $p < .001$, whereas the non-linear quadratic term did not approach significance, $t = 0.35$, $p = .34$. Consequently, we calculated residual reading times using only a linear effect of word length.

[7] The relation of log milliseconds to milliseconds is nonlinear, but at mean reading time, a difference of 1 unit of log residual reading time corresponds to a difference of 825 ms of residual reading time.

Table 3
Fixed effect estimates for multi-level logit model of memory judgments in Experiment 2 (N = 3368, log-likelihood = −1687).

Fixed effect	β	SE	Wald z	p
Baseline rate of *true* responses (response bias)	−0.33	0.11	−2.92	<.01
Rejections of unmentioned probe versus baseline (sensitivity)	3.32	0.25	13.20	<.001
Rejections of alternative versus baseline (sensitivity)	−0.11	0.31	−0.37	.71
Emphasized word (effect on response bias)	−0.03	0.09	−0.32	.75
Reading time (effect on response bias)	0.19	0.09	2.09	.04
Emphasized word × reading time (effect on response bias)	0.04	0.18	0.22	.83
Emphasized × rejections of unmentioned probe (effect on sensitivity)	−0.14	0.27	−0.53	.60
Emphasized × rejections of alternative probe (effect on sensitivity)	0.49	0.25	1.97	<.05
Reading time × rejections of unmentioned probe (effect on sensitivity)	−0.27	0.27	−0.99	.32
Reading time × rejections of alternative probe (effect on sensitivity)	−0.12	0.24	−0.48	.63
Emphasized × reading time × rejections of unmentioned probe (effect on sensitivity)	−0.27	0.54	−0.49	.62
Emphasized × reading time × rejections of. alternative probe (effect on sensitivity)	0.99	0.48	2.05	<.05

Note. SE = standard error.

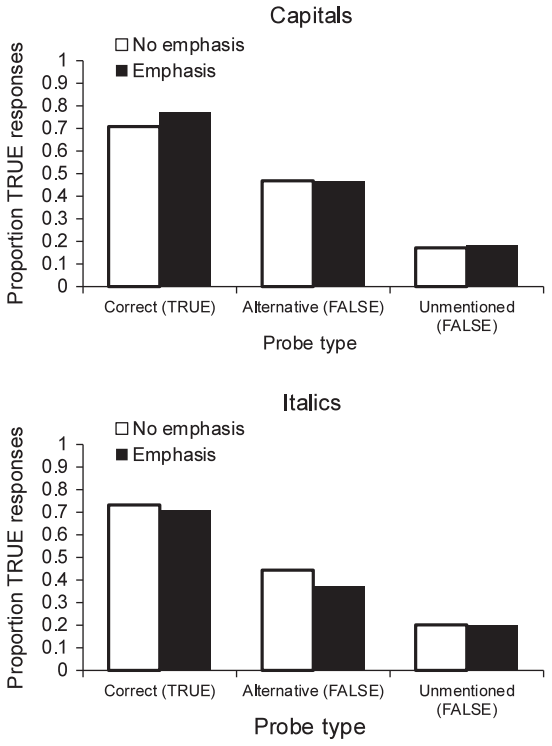

Fig. 3. Mean rate of *true* responses in Experiment 2 as a function of font emphasis and probe type, for participants who saw capitalization (top panel) and for participants who saw italicization (bottom panel). Responding *true* is a hit to a correct probe and a false alarm to an alternative or unmentioned probe.

in the opposite direction.) Increased reading time only improved memory when it was spent on emphasized words.

There was one effect on response bias: the odds of a *true* response were 1.21 times greater (95% CI: [0.98,1.38]) for every 1-unit increase in residual reading time, regardless of whether a particular probe was true or false, $z = 2.09$, $p < .05$. It is not clear what would account for this effect and it was not replicated in Experiment 3.

Discussion

Experiment 2 replicated the effects of font emphasis on memory observed in Experiment 1. Emphasizing one member of a pair using a font change enhanced participants' later ability to reject a proposition referring to the other, alternative member of the pair. But, it did not help reject propositions about items that were unmentioned, which were unlikely to be part of the alternative set. Experiment 2 generalized this pattern to new discourses different from the ones used in the original Fraundorf et al. (2010) study.

Experiment 2 extended the first experiment by collecting a measure of online reading time. Words emphasized with capital letters or italics were read more slowly than non-emphasized words. Moreover, the degree to which reading was slowed predicted the extent to which font emphasis helped reject the alternative probes on the later memory test. This relationship was observed even in a measure of residual reading time that removes any confound with participant-level differences in baseline reading speed. Moreover, the words and syntactic structures used in each discourse were invariant. Thus, the relation of reading time to later memory cannot be attributed to differences in lexical or syntactic variables that might affect reading time, such as word frequency, imageability, or syntactic position. Rather, this pattern might suggest that it is time-consuming to calculate an alternative set in response to emphasized words: the degree to which readers spent extra time on the emphasized words predicted the extent to later memory showed a benefit of encoding the alternative set.

Experiment 2 also clarified the difference between the two types of font emphasis tested in the present study: capitalization and italics. In Experiment 1, capital letters were observed to have a stronger mnemonic benefit than italics. In Experiment 2, capitalization also led to greater increases in online reading time; however, once these differences in initial reading time were controlled for, the two types of font emphasis did not differ in their effects on memory. That is, the difference between capitalization and italicization in their effects on memory appears to stem from how strongly they affected initial reading time.

The experiments thus far provide evidence that font emphasis can prompt readers to encode an alternative set, helping them to later reject certain salient alternative propositions. However, it is unclear exactly *which* alternative propositions are encoded in this set. One hypothesis is that the set of alternatives includes propositions about *any* item in the discourse of the same semantic or syntactic category—as might be expected if its contents are strongly influenced by generic or taxonomic knowledge. Another hypothesis is that the situation model tightly circumscribes the alternative set, as it does the set of items considered in resolving a referring noun phrase (e.g., Brown-Schmidt & Tanenhaus, 2008; Nieuwland et al., 2007). Under this latter hypothesis, comprehenders encode as alternatives only those propositions that are plausible or likely given the scenario described in the discourse. Either of these hypotheses could explain the results of Experiments 1 and 2: the alternative probes were both mentioned and good alternatives for the true proposition whereas the unmentioned probes involved items that were neither salient alternatives nor mentioned in the discourse at all. We conducted Experiment 3 to tease apart these hypotheses.

Experiment 3

In Experiment 3, we tested whether font emphasis would help reject propositions about items that were mentioned in the discourse but that were unlikely alternatives in a situation model. For example, consider context passage (6a), reproduced below.[8] Both *Saturn* and *Neptune* are mentioned exactly once in the discourse. But, they differ in their relationship to *Jupiter* in the continuation (6b). As in prior experiments, *Saturn* is mentioned as part of the same pair as *Jupiter* and is a plausible alternative for *Jupiter* in the continuation. (That is, the photos of Saturn could have been lost instead.) However, *Neptune* is a poor alternative for *Jupiter* in the situation model in (6b) because the discourse establishes that the mission to Neptune has not yet occurred.

(6a) Originally, the space probe Cosmo III was designed to fly past Jupiter and Saturn and send photos and measurements back to NASA from both planets. NASA needed this information to guide the videos they were going to take of Neptune on a future mission.

(6b) However, due to a glitch in the programming of the Cosmo III, it lost the photos taken of Jupiter and put the future mission in trouble.

Consequently, memory probes (8a) and (8b), although both false and both mentioning prior discourse entities, could be differentially affected by font emphasis. Probe (8a) represents the alternative probe condition, the same as in prior experiments, whereas probe (8b) involves an item that, while mentioned in the discourse, is a poor alternative in the situation model for the true proposition. We term this new probe condition the *merely mentioned probe*. If prominence leads comprehenders to encode only those alternative propositions that are plausible in a situation model, then font emphasis should help discriminate the correct probe only from the alternative probe and not from the merely mentioned probe. However, if the alternative set includes propositions involving any referent in the discourse of the same semantic category, then font emphasis should help discriminate the truth from both probe types.

(8a) NASA lost some of the data from Saturn due to a glitch in the space probe.

(8b) NASA lost some of the data from Neptune due to a glitch in the space probe.

Experiment 3 also eliminated the confound between probe conditions and lexical properties. In Experiments 1 and 2, the alternative and unmentioned probe conditions contained different lexical items. For example, for one discourse, the unmentioned probe was always *Portuguese*, while the alternative was *British* or *French*. It is possible that the sets of lexical items used in alternative probes versus unmentioned probes coincidentally differed in some property, such as lexical frequency, imageability, or general semantic plausibility, and that the differences between probe types were actually driven by these lexical differences rather than their relevance as an alternative to the true proposition. In Experiment 3, the same lexical items were rotated through the alternative and merely mentioned probe conditions across lists, thus controlling for any lexical properties that might have influenced the effect of prominence.

As in Experiment 2, we collected measures of reading time in Experiment 3 to further test whether variability in the influence of font emphasis on memory was related to whether readers invested additional time in processing the emphasized words.

Method

Participants

Forty-eight new participants completed Experiment 3.

Materials

Thirty-six discourses were constructed for Experiment 3. The discourses took the same general format as those used in previous experiments. In Experiment 3, however, each context passage introduced not only two pairs of items, but also one additional item related to each pair. This third item was from the same semantic domain, but the discourse established that it was not participating in an event, had occurred or would occur at a different time, or had already been ruled out by a decision-maker, thus making it an unlikely alternative in the situation model for the item mention in the continuation passage. The alternative and the merely mentioned item were each mentioned exactly once in the discourse. For example, in example (6) above, the context passage establishes the pair

[8] This discourse contained the capitalized acronym *NASA*. This was the only such item that contained an acronym; otherwise, words were only written in all capitals when they were part of the manipulation of font emphasis.

Jupiter and *Saturn*; *Neptune* is also mentioned, but in a context that establishes it is not part of the mission described in the discourse. The same applies for the pair *photos* and *measurements* versus the merely mentioned item *video*.

As in Experiment 2, the target words in the continuation never appeared at the beginning or end of a line of text on the screen and never at the beginning or end of a punctuation-marked clause. Also as in prior experiments, an equal number of continuations referred to the member of the pair that the context passage had mentioned first as had been mentioned second.

The correct probe and the alternative probes for the recognition memory test were constructed as in previous experiments. The unmentioned probes were replaced by the merely mentioned probes, which referred to the item that the discourse had established as an unlikely alternative to the true proposition (e.g., *Neptune* in the above example).

In Experiment 3, the lexical items used for the alternative and merely mentioned probes were rotated across lists. That is, one participant would see the pair *Jupiter and Saturn* with *Neptune* as the merely mentioned item, while another participant would see *Jupiter and Neptune* with *Saturn* as the merely mentioned item. As our primary interest was in the two types of false probes from which the correct information needed to be discriminated, the true proposition itself was consistent across lists.

The rotation of lexical items across lists introduced additional counterbalancing variables. To avoid a proliferation of experimental lists, we did not manipulate the type of font emphasis used in Experiment 3. For all participants, emphasized words were emphasized with capital letters. Prior experiments had observed qualitatively similar patterns across emphasis types, especially when initial reading time was taken into account.

The Experiment 3 discourses appear in Appendix C and the test probes in Appendix D.

Procedure

Aside from the change in materials, the procedure of Experiment 3 was identical to that of Experiment 2.

Results

Initial reading time

Reading times during the initial presentation of the discourse in Experiment 3 are displayed in Fig. 4 as a function of region and presence of font emphasis. We analyzed the reading times on the critical word and spillover region as in Experiment 2; as before, reading times were highly skewed (skewness = 11.16), so we used the natural log of the reading times (skewness = 0.93).

Random slopes by subjects for the two factors significantly improved the fit of the model, $\chi^2_{(9)} = 239.02$, $p < .001$, as did the same slopes by items, $\chi^2_{(9)} = 83.67$, $p < .001$. Fixed effect parameter estimates from the full model are displayed in Table 4. Again, words with font emphasis ($M = 510$ ms) were read more slowly than the same words without emphasis ($M = 417$ ms), $t = 5.47$, $p < .001$. Reading times did not significantly differ between the critical and spillover words, nor did the effect of font emphasis reliably vary across these regions.

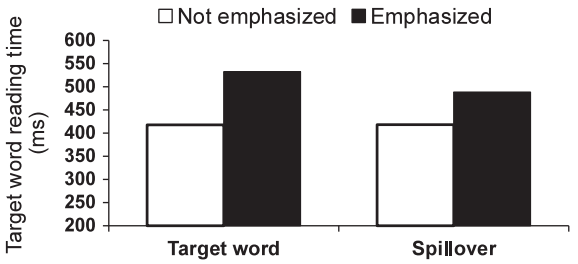

Fig. 4. Mean reading time in Experiment 3 on target words and spillover words as a function of font emphasis.

Table 4
Fixed effect estimates for multi-level model of log reading time in Experiment 3 ($N = 6912$, log-likelihood: −3281).

Fixed effect	β	*SE*	*t*	*p*
Non-emphasized critical word (baseline)	5.91	0.05	129.87	<.001
Emphasized word	0.15	0.03	5.47	<.001
Spillover region	<0.01	0.02	0.41	.68
Spillover region × emphasized word	−0.02	0.03	−0.79	.43

Note. SE = standard error.

Memory

Proportions of *true* responses are displayed in Fig. 5 as a function of probe type and presence of font emphasis in the original discourse. As in Experiment 2, to assess whether later memory performance was predicted by online reading of the initial discourse, we included as a predictor the residual reading time calculated from the log-transformed reading times. Residual reading times were summed over the critical and spillover words, as both showed sensitivity to the font emphasis (as demonstrated above). Observations with reading times more than three standard deviations from a participant's mean in a condition were eliminated, affecting less than 1% of the data.

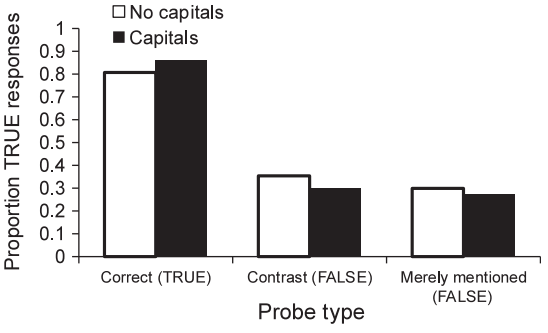

Fig. 5. Mean rate of *true* responses in Experiment 3 as a function of font emphasis and probe type. Responding *true* is a hit to a correct probe and a false alarm to an alternative or merely mentioned probe.

We then modeled recognition memory decisions as a function of probe type, font emphasis, and residual reading time, as well as the random effects of participants and items. Once again, the model fit was improved by a random slope of probe type by participants, $\chi^2_{(5)} = 51.40$, $p < .001$, and by items, $\chi^2_{(5)} = 86.98$, $p < .001$. No further slopes contributed reliably to the model, all p values >.10. Fixed effect parameter estimates for the final model are displayed in Table 5.

Overall, participants showed sensitivity to the truth value of the probes. The odds of rejecting a merely-mentioned probe were 8.70 times greater (95% CI: [5.72,13.24]) than the baseline rejection rate (the first planned comparison), $z = 10.11$, $p < .001$, and the odds of rejecting an alternative probe were 5.45 times greater (95% CI: [3.71,8.02]) than the baseline (the second planned comparison), $z = 8.62$, $p < .001$.

Font emphasis improved rejections of the alternative probes, as in previous experiments, with the odds ratio between rejections of alternative probes and the overall rejection rate being 1.85 times (95% CI: [1.15,2.97]) greater for emphasized words than for non-emphasized words, $z = 2.53$, $p < .05$. However, font emphasis did not reliably benefit rejections of merely mentioned probes. Although the odds of those rejections were 1.29 times (95% CI: [0.850,2.10]) greater with font emphasis, the effect was not statistically reliable, $z = 1.04$, $p = .30$. Moreover, the model parameter estimates indicate that the effect of font emphasis was over twice as large for rejecting alternative probes (an increase of 0.61 in the log odds) than for rejecting merely mentioned probes (an increase of 0.26).

This pattern is what would be expected if font emphasis only benefited rejections of the most salient alternatives and not about a mentioned but implausible alternative. However, one other account must be considered. In many of the discourses, such as item (6) above, the merely mentioned item was mentioned in a different sentence than the correct item and salient alternative item, and it is possible this difference drove the difference between probe types. To assess whether mention in the same sentence accounted for the results, we conducted a post hoc coding of whether the merely mentioned item was mentioned in the same sentence as the salient alternative. This factor and its interactions with the other variables of interest were then entered into the model. This new variable and its interactions did not significantly contribute to the overall fit of the model, $\chi^2_{(12)} = 10.40$, $p = .58$, nor did mention in the same sentence participate in a significant three-way interaction with either of the critical font emphasis x probe type interactions (both p values >.40). Thus, the data do not provide any support for the hypothesis that the difference between the alternative and merely mentioned items was driven by which sentence they were mentioned in.

Unlike in Experiment 2, initial reading time had no effect on any aspects of participants' responding in the recognition memory task. In fact, increased reading time on emphasized words numerically predicted *worse* discrimination between correct and alternative probes, although this effect did not quite reach conventional levels of significance ($p = .05$).

Discussion

Experiment 3 replicated the finding that font emphasis benefits recognition memory by improving participants' ability to reject a salient alternative to a true proposition. Experiment 3 also introduced a new type of memory probe that referred to an item mentioned in the discourse but that the discourse had established as a poor alternative to the true proposition. Font emphasis conferred no benefit in rejecting these probes. These results suggest that the font emphasis led readers to encode only those alternative propositions that were plausible given the situation described in the discourse. Had every discourse entity of the same semantic or syntactic category been encoded in an alternative proposition, font emphasis should have benefited rejection of both lure types. Instead, the font emphasis benefited only rejections of a particularly important alternative to the true proposition.

The experiments were inconsistent in whether they showed a relation between reading time and memory. In Experiment 2, longer reading time on emphasized words predicted superior rejections of salient alternatives, suggesting the extra time was being used to encode the alternative set. Experiment 3 failed to replicate this relation; indeed, reading time did not reliably predict any aspect of performance on the recognition memory task. There

Table 5
Fixed effect estimates for multi-level logit model of memory judgments in Experiment 3 ($N = 3422$, log-likelihood = −1785).

Fixed effect	β	*SE*	Wald z	p
Baseline rate of *true* responses (response bias)	−0.06	0.09	−0.67	.50
Rejections of merely mentioned probes versus baseline (sensitivity)	2.16	0.21	10.11	<.001
Rejections of alternative probes versus baseline (sensitivity)	1.70	0.20	8.62	<.001
Emphasized word (effect on response bias)	>−0.01	0.09	>−0.01	.99
Reading time (effect on response bias)	−0.01	0.08	−0.09	.93
Emphasized word × reading time (effect on response bias)	−0.14	0.15	−0.89	.38
Emphasis × rejections of merely mentioned probes (effect on sensitivity)	0.26	0.25	1.04	.30
Emphasis × rejections of alternative probes (effect on sensitivity)	0.61	0.24	2.53	<.05
Reading time × rejections of merely mentioned probes (effect on sensitivity)	0.27	0.22	1.27	.21
Reading time × rejections of alternative probes (effect on sensitivity)	−0.21	0.21	−0.99	.32
Emphasis × reading time × rejections of merely mentioned probes (effect on sensitivity)	0.41	0.43	0.96	.34
Emphasis × reading time × rejections of alternative probes (effect on sensitivity)	−0.81	0.42	−1.92	.05

Note. SE = standard error.

are two possible reasons why reading time on emphasized words was unrelated to discriminating correct and alternative probes in Experiment 3. One reason is that calculating and encoding the alternative set did not require any additional time. (Although readers did read the emphasized words more slowly even in Experiment 3, this additional time could have been spent on processes such as decoding the capital letters rather than encoding an alternative set.) The other potential reason is that encoding the alternative set did require extra time but that readers did so every time they encountered an emphasized word. This behavior would also result in a null relation between memory and reading time on emphasized words: if *every* encounter with an emphasized word was one in which readers slowed down to encode the alternative set, there would not enough be variability in participants' reading times for them to correlate with later memory.

Experiment 3 succeeded in eliminating the baseline differences between the types of false probes. The alternative probes (M = 33% affirmed) and merely mentioned probes (M = 29% affirmed) were affirmed at similar rates overall. Nevertheless, only rejections of the alternative probe benefited from the font emphasis; rejections of the merely mentioned probe did not. This pattern provides strong evidence that font emphasis led to encoding of only particularly plausible alternatives and that the difference between probe types does not simply reflect a floor effect.

The comparison between the alternative and merely mentioned probes also rules out two other interpretations of the results of Experiments 1 and 2. One alternate account of the first two experiments is that font emphasis simply served to strengthen the overall representation of the correct discourse, but that the consequences of this enhancement for responding on the recognition memory test were modulated by whether or not the test probes reactivated the original discourse. Because both the correct and alternative probes included referents mentioned in the original discourse, they could have reactivated the original discourse in memory and allowed rememberers to reap the benefits of a representation that had been strengthened by font emphasis. By comparison, the unmentioned probes centered on items that had never been mentioned in the discourse; if these probes thus failed to reactivate the original discourse representation, it would not matter whether font emphasis had strengthened the discourse representation. However, Experiment 3 provides evidence against this reactivation account because the merely mentioned probes referred to items that *had* been mentioned in the original discourse. If the key difference between probe conditions was simply whether they reactivated the original discourse, then font emphasis should also have modulated responding to the merely mentioned probes. However, no such effect was observed even when, as noted above, the merely mentioned items appeared in the same sentence as the correct and alternative item. Moreover, it is not clear that the unmentioned probes in the first two experiments would have failed to reactivate the discourse. Even the unmentioned probes included many other words from the original discourse, such as *scientists*, *endangered*, and *monkey* in example (4), that could have reactivated the discourse representation. Thus, it appears the critical difference between probe types is in the discourse role of the probed words, not simply whether the items had appeared in the original discourse and could reactivate the discourse.

Experiment 3 also rules out an account based on familiarity. It has been proposed that recognition memory decisions may be made in part simply on the bases of familiarity or recency of the probes (cf., Yonelinas, 2002). In Experiments 1 and 2, the alternative probes referred to items that had been mentioned and were more familiar in the context of the experiment, whereas the unmentioned probes referred to items that had not been recently mentioned at all and, as a consequence, were also less familiar. Thus, it was possible that the differential effects of font emphasis on contrast versus unmentioned probes in prior experiments reflected a difference between effects on familiar versus unfamiliar words rather than between items inside versus outside an alternative set. However, in Experiment 3, both the alternative and merely mentioned probes referred to items that were mentioned and familiar within the context of the experiment. Nevertheless, font emphasis selectively benefited only discrimination between the truth and the alternative and not discrimination between the truth and the merely mentioned item, suggesting that its benefit lay in encoding of that alternative in an alternative set.

Discussion

Three experiments tested the effect of font emphasis on later memory for a discourse. In all three experiments, font emphasis during the original presentation of a discourse promoted the encoding of an alternative to a true proposition. This is evidenced by the selective mnemonic benefit of emphasizing one member of a pair of words: emphasis helped participants reject a salient alternative proposition, but it did not benefit rejections of other kinds of false statements. This pattern is consistent with past results from contrastive (L + H*) pitch accents (Fraundorf et al., 2010), but the fact that it was now observed in written discourse as a function of font manipulations demonstrates that it is not idiosyncratic to explicit manipulations of prosody. Moreover, Experiment 3 provided evidence that readers encode only those alternatives that are plausible given the situation described by the discourse: font emphasis helped reject those plausible alternatives from the truth, but it did not help reject propositions about entities that were mentioned but implausible alternatives in the situation model.

The relation of online reading time in modulating these representations was investigated in Experiments 2 and 3, but the results were inconsistent. In Experiment 2, increased reading time on emphasized words amplified their benefit in rejecting the salient alternatives, suggesting that an alternative set was more likely to be encoded, or was better encoded, with additional time. In Experiment 3, however, no relations of reading time to memory performance were observed. Given the similarity of the experiments in materials and procedures, it is unclear what would account for these discrepant findings. Mean reading times were roughly similar across Experiment 2

(M = 856 ms for the critical word and spillover region combined) and Experiment 3 (M = 901), as was the variability both between subjects (Experiment 2 SD = 276, Experiment 3 SD = 318) and within subjects (Experiment 2 SD = 362, Experiment 3 SD = 332). One possibility is that the reading times collected in the moving window task, although generally strongly correlated with reading times from more naturalistic tasks (Just et al., 1982), simply did not provide the most naturalistic assessment of how readers read emphasized text and may have obscured the relation of reading time to comprehension. We also note that past investigations of the relation between online reading time and offline discourse comprehension have produced similarly mixed results (e.g., Caplan et al., 2011; Christianson & Luke, 2011; Daneman et al., 2007; Reder & Kusbit, 1991; Ward & Sturt, 2007).

Emphasized text and memory

In all three experiments, emphasizing words with capitalization or italicization led to superior performance on a recognition memory test—in particular, by helping readers to reject a salient, but false, alternative proposition. This pattern suggests that, at least given a salient alternative present in the discourse, the memory benefit of font emphasis stems from encoding a set of alternatives to a true proposition (e.g., Calhoun, 2009; Rooth, 1992; Steedman, 2000). The effect of the font emphasis was not simply to change how the emphasized referent itself was encoded. Rather, font emphasis altered readers' representations of *other* elements of the discourse by leading readers to encode them in alternative propositions to what actually happened in the discourse.

A similar benefit to memory has been produced by contrastive (L + H*) pitch accents (Fraundorf et al., 2010). Evoking contrast has been argued to be the meaning of the L + H* intonational contour (Gussenhoven, 1983; Pierrehumbert & Hirschberg, 1990), and it was possible that other linguistic devices might not carry the same contrastive meaning and thus not produce the same memory benefits. The present results, however, suggest that the effect generalizes beyond an explicit presentation of the L + H* pitch contour because it also obtained from a manipulation of font emphasis in written discourse.

Why does font emphasis produce similar effects as contrastive (L + H*) pitch accents? Because readers generate prosody even when reading silently (e.g., Breen & Clifton, 2010; Fodor, 1998), they may have generated the L + H* accent in their implicit prosody when they encountered the emphasized text. Another explanation comes from recent proposals (Calhoun, 2009) that *any* linguistic element that is more prominent than expected can promote the consideration of salient alternatives. In this account, font emphasis could lead to consideration of a salient alternative even without implicitly evoking an L + H* pitch accent simply because it makes words more prominent. It will be necessary for future research to directly compare these hypotheses, perhaps by attempting to eliminate implicit prosody through articulatory suppression (Besner, 1987). Nevertheless, the present experiments show that the representation of salient alternatives in discourse does not depend on an overt presentation of a contrastive L + H* pitch accent.

The results of the present experiments are also noteworthy for demonstrating clear memory benefits from font emphasis. Prior results have been mixed as to whether font emphasis actually improves learning from a text (for review, see, Hartley, Bartlett, and Branthwaite (1980)). The present data suggest that one reason for these mixed results may be that font emphasis can have a relatively nuanced effect on comprehension, helping readers to encode only certain kinds of information and benefiting only certain kinds of mnemonic decisions. Consistent with this, Golding and Fowler (1992) found that underlined words—in concert with several other manipulations designed to make important details more prominent—improved learning of a text only when readers were told that they would be tested on specific details rather than on the overall gist.

Although font emphasis benefited memory in all three experiments, the magnitude of its benefit varied somewhat across experiments: font emphasis increased the log odds of correct rejections twice as much in Experiment 1 as it did in the other experiments. One potential reason for the larger difference between emphasized and non-emphasized words in Experiment 1 is that the entire discourse was visible to participants at one time, increasing the salience of the distinction between the emphasized word and the rest of the text. Indeed, readers in Experiment 1 could have adopted a strategy of skimming for the emphasized words, as they might do when reading a textbook, whereas this strategy would not have been possible when word-by-word reading was enforced in Experiments 2 and 3. Nevertheless, font emphasis benefited memory in all three experiments, demonstrating that its effects are not dependent on a particular mode of presentation.

What constitutes an alternative set?

The present experiments also tested what constrains the set of alternatives that is encoded in response to prominence. Although it is generally agreed (e.g., Rooth, 1992; Van Deemter, 1999) that the set of alternatives must be constrained by context, it has been unclear which aspects of the context influence the set. One hypothesis was that mere mention in the context is sufficient, such that all items in the same semantic or syntactic category would be incorporated into propositions in the alternative set. This pattern might be predicted from accounts in which consideration of alternatives is strongly influenced by generic or taxonomic knowledge (e.g., Blok & Eberle, 1999) and from findings that givenness is a strong influence on discourse processing (e.g., Bock & Mazzella, 1983; Halliday, 1967; Haviland & Clark, 1974). A different hypothesis was that the alternative set is constrained by the particular situation described by the discourse (Zwaan & Radvansky, 1998), such that it only included propositions about those entities that were particularly plausible alternatives given the situation model. The present experiments supported the latter hypothesis. In Experiments 1 and 2, font emphasis helped readers reject a salient alternative proposition, but it did not help them reject a proposition about an item

that was from the same semantic category but that was unmentioned in the discourse. In Experiment 3, an even stricter test, font emphasis did not help reject a proposition about an item that was mentioned in the discourse, but in a context that made it a less plausible alternative in the situation model. These results imply that comprehenders consider and encode only those alternatives that are plausible given the situation model—a relatively narrow set—in the alternative set, rather than a larger set incorporating all mentioned items.

This result is consistent with results from other linguistic domains indicating that the situation described by a particular discourse can tightly constrain the number of relevant alternatives. Seemingly ambiguous referring expressions can be interpreted without difficulty if some of the candidate referents are physically distant or task-irrelevant (Brown-Schmidt & Tanenhaus) or have exited the scene described by a text (Nieuwland et al., 2007), thus placing them outside the domain of reference. Moreover, the discourse can add alternatives from outside a semantic category. For example, reading that *Christina wants to buy a lock, nails, and a bolt* can make *nails* an alternative for *lock* even though they belong to different semantic categories (Byram Washburn et al., 2011). This evidence all indicates that the situation model, and not only generic categorical knowledge, constrains what set of alternatives is considered.

Importantly, the results of Experiment 3 also demonstrate that the differential effects of font emphasis on the salient alternatives versus other false probes cannot be attributed simply to familiarity, recency, or prior mention in the discourse. In Experiment 3, both the alternative probes and the merely mentioned probes had been mentioned in the same discourse, but font emphasis helped discriminate only the alternative probes from the truth. Thus, the locus of the emphasis benefit was in distinguishing the truth from particularly salient alternatives to the true proposition.

Perceptual characteristics of emphasized text

Although the present experiments suggest that font emphasis led readers to encode a salient alternative to the outcome of the discourse, font manipulations also alter the perceptual properties of text. One alternate explanation might thus be that font changes simply made words more distinct, which is known to enhance memorability (the von Restorff effect; e.g., Hunt & Lamb, 2001). In addition, words with unusual fonts may read more slowly or with increased difficulty. This feeling of disfluency has sometimes been argued to promote deeper, more effective processing (Diemand-Yauman, Oppenheimer, & Vaughan, 2011).

However, several facts suggest the present data cannot be attributed solely to the perceptual characteristics of the words. First, the pattern of memory performance indicates that the effect of font emphasis was not to modulate the representation of the target word but of the salient alternative proposition. If the prominence of *British* led to enhanced representation of the word *British* itself, that should have helped reject all false statements, all of which would be inconsistent with the enhanced representation of *British* (Brainerd et al., 2006). Instead, emphasizing *British* only helped reject a particularly relevant alternative, not probes referring to items that were unmentioned or that were mentioned as less plausible alternatives. This implies the effect of the font manipulation did not lie in the representation of *British* itself. Second, the effect of the font manipulation is likely to have arisen not from mere perceptual difficulty: adding random visual noise to text impairs, rather than benefits, comprehension (Gao, Stine-Morrow, Noh, & Eskew, 2011). Rather, the benefit was likely in the interpretation that readers gave the font.

Nevertheless, it is clear that the perceptual characteristics of prominent words play a role in how they are processed. Perceptual distinctiveness may be a cue that additional processing is needed. In the presence of a relevant alternative, however, it appears the effect of this prominence is not to globally enhance memory but to promote encoding of the alternative. This is consistent with the view (Calhoun, 2009) that prominence is one device by which speakers or writers can deliberately signal that salient alternatives should be considered.

Language comprehension and long-term memory

The present study also highlights the importance of linguistic context in constraining memory. Perceptual distinctiveness (e.g., Hunt & Lamb, 2001) and longer study time (e.g., Dunlosky & Connor, 1997) are often associated with generally superior memory. In the present experiments, however, font emphasis helped only certain kinds of mnemonic decisions: those that involved rejecting a salient alternative proposition suggested by the discourse. Similarly, the mnemonic benefits of additional reading time observed in Experiment 2 applied only to this discrimination.

These results join other recent findings demonstrating that linguistic context can constrain or override other mnemonic effects. For example, reading words in a sentence context as opposed to a word list increases erroneous endorsement of semantically related lures, such as *nosedive* for *tailspin,* but decreases endorsement of morphemically related lures, such as *tailgate* for *tailspin* (Matzen & Benjamin, 2009). Linguistic context also modulates the effects of repetition. Typically, repeated words are easier to process and remember. However, when the discourse context makes a repeated word less felicitous than a pronoun, lexical repetition impairs processing and memory (Almor & Eimas, 2008; Ledoux, Gordon, Camblin, & Swaab, 2007).

The role of contrast in cognition

The present study focused on whether representing salient alternatives contributed to long-term, offline memory for a discourse, but there is also evidence that knowledge about alternatives contributes to initial, online language processing. For example, scalar adjectives such as *small* are usually used to select items from an alternative set; that is, distinguishing a *small glass* from a set of other glasses that are not small. Consequently, when listeners hear the beginning of a noun phrase containing a scalar adjective, such as *the small–*, they preferentially look to

an object that has an alternative of a different size (e.g., a small glass when there is also a large glass present) rather than a singleton object without an alternative (e.g., a small bowl with no large bowl present; Sedivy, Tanenhaus, Chambers, & Carlson, 1999). This looking preference obtains even if the alternative is no longer visually copresent (Wolter, Gorman, & Tanenhaus, 2011), providing further evidence that salient alternatives are maintained in memory. Further, it is not only scalar adjectives that favor a contrastive interpretation in online processing. Listeners look to an object that has an alternative of a different color when a contrastive (L + H*) pitch accent is placed on a color adjective (Ito & Speer, 2008). And, a bare noun phrase with an L + H* accent is preferentially interpreted as referring to an object that has a visually copresent alternative with a similar sounding name (e.g., a *camel* when a *candle* is also present; Watson, Gunlogson, & Tanenhaus, 2008).

The computation of alternatives has been observed in other domains as well. For example, novel categories may be learned in part by contrasting them with salient alternative categories, and doing so may exaggerate the differences between the categories in memory (Davis & Love, 2010). The consideration of alternatives may be a robust component of cognition; it is observed across a variety of domains and as a function of a variety of manipulations (McAteer 1992; Sanford, Price, & Sanford, 2009; Sedivy, 2002; Sedivy et al., 1999).

Conclusion

Font emphasis, just like some prosodic pitch accents, influences discourse representations not by changing how true propositions are encoded but by leading readers to represent alternative propositions in a discourse. The set of alternatives appear to be constrained by the situation model, with font emphasis helping rule out only those alternatives that are plausible in the scenario described by the discourse. Knowledge of these alternatives can benefit later memory by allowing what actually happened to be more easily distinguished from other likely possibilities. Thus, the present experiments reinforce a view in which successful memory for the events of a discourse is supported by encoding not only what happened but also what did *not* happen.

Acknowledgments

Scott H. Fraundorf, University of Illinois at Urbana–Champaign; Aaron S. Benjamin; University of Illinois at Urbana–Champaign; Duane G. Watson, University of Illinois at Urbana–Champaign.

This work was supported by National Science Foundation Graduate Research Fellowship 2007053221 and a traineeship from National Institutes of Health Language Processing Training Program 5T32HD055272-13 to Scott H. Fraundorf and by National Institutes of Health grants R01AG026263 to Aaron S. Benjamin and R01DC008774 to Duane G. Watson. The experiments reported here served as part of the first author's dissertation. We thank Sarah Brown-Schmidt, Gary Dell, Kara Federmeier, Elizabeth Stine-Morrow, and the members of the Communication and Language Lab and the Human Memory and Cognition Lab for their comments and suggestions, Melissa Denis, Ran He, Euil Lim, Gina Mulanthara, and Sujin Park for data collection, and Martha Fraundorf for assistance with stimulus construction.

Appendix A

Stimuli for Experiment 2

1. *Context:* Steve and his wife had been wanting to visit the Badlands and the Redwoods. Steve could take his vacation either in the spring or in the summer.
   Continuation: Steve and his family eventually decided to visit the Redwoods in the spring because they found a good deal on a hotel.
2. *Context:* Annette completed all her holiday shopping for her father and her nephew while on vacation in Spain. She stopped in a gift shop in Barcelona and bought a shirt and some books of Spanish art.
   Continuation: She gave her nephew the books for Christmas. He was very happy with the gift and said it was his favorite of the gifts he received that year.
3. *Context:* The local parks commission had a busy meeting on Wednesday to decide how to spend its money for the year. People disagreed on whether the commission should focus its resources on expanding the parks or repairing the existing parkland. The parents wanted to add a playground area but the teenagers wanted a skateboard area.
   Continuation: After a long debate, a compromise was worked out to repair the parks and build a skateboard area, but it didn't seem like anyone was very happy with the decision.
4. *Context:* Two best friends from college, Matt and Eric, ended up as rival salesmen selling Toyota and Subaru cars.
   Continuation: They had a friendly competition going, but most years Matt was the winner because of his Toyota deals. He got to keep the silly trophy they'd made for the contest.
5. *Context:* After the old mansion was finally sold to a new owner, an inspector was sent in to check for pests and leaks in both the bathroom and basement.
   Continuation: He discovered a major pest problem in the mansion's basement and said a specialist would be needed to fix it. It was one of the worst messes he'd ever seen.
6. *Context:* As soon as the word about the burglary and the accident reached the newspapers, photographers from the Gazette and the Journal rushed to the scene to cover the stories.
   Continuation: The front-page photo of the

(*continued on next page*)

accident scene in the Journal later won an award.

7. *Context:* Andrea never enjoyed flying, but with both a professional meeting and a wedding to attend this month, she knew she'd just have to endure it. She only hoped her flights weren't canceled or rerouted.
   Continuation: Unfortunately, her fears came true when her flight to the wedding was canceled with very little notice. Andrea wrote an angry letter to the airline about the situation.
8. *Context:* When the rock band first formed, the bandleader handled both guitar and vocals. Later, after Laura and Chris joined the band,
   Continuation: Laura took over on vocals for the group.
9. *Context:* The director and producer of a forthcoming film were having a big dispute over the casting and special effects.
   Continuation: They failed to reach an agreement on the casting and the director quit in a rage.
10. *Context:* Both the teachers and the students were pleased with the renovation of the school, which made the rooms larger and warmer during the cold winter months.
    Continuation: The teachers were particularly pleased to have the warmer rooms. But, everyone thought the new carpet was really ugly.
11. *Context:* The bride and the groom disagreed about whether the wedding should be held at a chapel or at a resort.
    Continuation: But the bride had a connection that helped them reserve a resort for a Valentine's Day wedding, so that is what they chose.
12. *Context:* Some ecologists went to Africa to check on the rhino and hippo populations in Kenya and Sudan.
    Continuation: They were happy to discover that the rhino herds in Kenya were getting larger.
13. *Context:* Although Jennifer owned both a cat and a dog, the two pets got along great with each other. There wasn't a problem until her cousin visited with her ferret and rabbit.
    Continuation: Jennifer's cat hated the cousin's ferret and chaos broke out in the house.
14. *Context:* Strangely enough, on the same day, both the Seattle-based publisher and the Houston-based publisher released new books about the American Revolution and about the Great Depression.
    Continuation: Critics judged the book on the Depression from the Seattle publisher to be the best of the bunch.
15. *Context:* Mike's doctor told him that he should get more exercise, so Mike considering walking or biking to work. He also thought about swimming or hiking.
    Continuation: But the only way he could fit all those activities into his schedule was to bike to work and swim on the weekends. After a few months, Mike's doctor was quite pleased with his progress.
16. *Context:* When Lindsay and Jessica saw each other at the high school reunion, they were surprised and amused by each other's career choices. Both of them had said they'd never want to be lawyers or bankers,
    Continuation: but, sure enough, Jessica was now working as a lawyer in Chicago. They both had a good laugh about it.
17. *Context:* To win the hand of the baroness's daughter, the German and the French knights competed in a tournament of fencing and archery.
    Continuation: Both knights gave it their best, but the French knight emerged victorious during the archery event and married the daughter.
18. *Context:* Samantha's biology and physics midterms were both the same week and she was so stressed that it was hard to study. She tried both the cafe and the quad but couldn't concentrate.
    Continuation: But, things started going better when three of her classmates invited her to a biology study group at the cafe and she felt more confident.
19. *Context:* Tina was disappointed when she realized that both the film and the play were scheduled for Friday evening, and that she could only go to one of them. She was really busy, so she thought she would go to the one that was shorter. But, then her boyfriend suggested the cheaper alternative to save money.
    Continuation: By Friday morning, she still hadn't decided, but she and her boyfriend eventually chose the play since it was the cheaper event. They enjoyed it and were glad they went.
20. *Context:* The local bowling lanes had gotten to be quite popular. The lanes were inexpensive and open to everyone of all skill levels, as long as they did not bring in any outside snacks or drinks. Thanks to these friendly policies, bowling leagues had been formed by both the professors and the architects.
    Continuation: But when architects were found with snacks that were not allowed, management had no choice but to temporarily ban their group from the lanes.
21. *Context:* The meteorologist had predicted bad weather for the 3-day weekend, saying that there might be rain or snow on Sunday and Monday. Julia thought about canceling her trip to the mountains when she heard the forecast, but she went anyway.
    Continuation: The only bad weather ended up being the rain that fell on Monday morning, so she was glad she didn't cancel her plans.
22. *Context:* Ben had made it to the last round of the game show. Now he had to choose to open either

the white door or the black door to claim his prize. Behind one of the doors was a boat and behind the other was a goat, but he only got one chance to pick.
Continuation: Nervously, Ben opened the black door and discovered the goat behind it. The audience went wild.

23. *Context:* Originally, the space probe was designed to fly past Jupiter and Neptune and send photos and videos back to NASA from both planets.
Continuation: However, due to a glitch in the system, the videos taken of Neptune were lost completely.
24. *Context:* The night before gameday, the quarterback always polished his rings and shoes for good luck. He thought that if he didn't, his team's rushing and passing wouldn't go as well.
Continuation: So, when the team blew a big passing play during the second quarter, he blamed it on the fact that he had forgotten to polish his shoes the night before.
25. *Context:* The renowned fashion designer divided his time between Milan and Paris while working on his new lines of shirts and skirts.
Continuation: His shirts were a hit when he revealed them at a Paris fashion show, but the others were less favorably received.
26. *Context:* A sporting goods manufacturer was looking for some new athlete endorsements for its lines of jackets and watches. Representatives from the company met with a swimmer and a baseball pitcher.
Continuation: The company signed the swimmer to endorse the watches in a commercial, but they dropped him after he became involved in a scandal.
27. *Context:* Elizabeth was in charge of organizing her friend Mary's birthday party. She asked Mary's brother if Mary would prefer lemon cake or spice cake and if she should serve punch or cider.
Continuation: Mary's brother suggested cider but didn't know about the cake. Elizabeth picked a spice cake, which she personally liked best.
28. *Context:* Bridget's niece was visiting and Bridget wanted to take her to either the history museum or the science museum. She wasn't sure about the traffic, so she checked online to see if it would be easier to go by bus or car.
Continuation: They could easily get to the history museum by car right as it opened.
29. *Context:* After the McKee Company struck business deals with firms in China and Russia, the company hired Suzanne as a translator.
Continuation: When the company was ready to send out letters and an invoice to one of its partners, Suzanne's first job was to translate the invoice from Russian as soon as possible.
30. *Context:* The juniors and seniors at the university were competing to raise the most money to fight hunger and cancer, so they held a number of fundraisers.
Continuation: The most successful was the haunted house, which helped the seniors raise even more money to fight cancer than they had last year.
31. *Context:* Jason's girlfriend invited him to go birdwatching at the river and marsh. Jason was expecting that it would be easy to spot the ducks and swans, but it was early in the morning and he was still sleepy. He didn't see most of the birds until they were pointed out to him.
Continuation: But he beat the rest of the group by spotting one of the ducks at the river first. He was excited and he told his girlfriend he'd be willing to go birdwatching again.
32. *Context:* The American and the Japanese engineers competed to make computer chips for phones. They kept working to make the chips smaller and quicker.
Continuation: The chip made by the American engineers was the most successful because it was the smallest of any available.
33. *Context:* A new Mexican and a new Italian restaurant had recently opened in the city. Both were waiting to hear whether or not the notoriously harsh food critic would give his approval to their specials and desserts.
Continuation: The critic originally planned to dine at both restaurants during the week, but because he caught the flu, he only had a chance to visit the Mexican restaurant, where he gave the desserts a favorable review.
34. *Context:* The small town was thrilled when two of its residents won Pulitzer Prizes for literature and journalism in the same year. The mayor planned to hold a parade and dinner in their honor.
Continuation: But, the winner of the prize for journalism told planners he hated publicity and would not attend the dinner under any circumstances. They held it anyway.
35. *Context:* A new children's movie has just been released about a farm where the pigs and cows can all talk. The farmers think that the crops are vulnerable to drought and disease.
Continuation: But when a severe drought strikes, a heroic pig ends up saving the farm from bankruptcy.

Appendix B

Test probes for Experiment 2

1A. Steve and his family decided to go to the (Redwoods/Badlands/Everglades) on vacation.

(*continued on next page*)

1B. Steve and his wife decided to visit a national park for vacation during the (spring/summer/fall).
2A. Annette gave her (nephew/father/brother) a gift from the gift shop in Barcelona that he was very happy to get.
2B. Annette's gift of (a shirt/books/a mug) from Barcelona was a favorite for its recipient.
3A. The local parks commission decided to (repair/expand/landscape) the parkland.
3B. The local parks commission decided to build a (skatepark/playground/dog park).
4A. Most years, the winner of the sales contest between college friends was (Matt/Eric/Nick).
4B. Most years, the winner of the sales contest between college friends was the (Subaru/Toyota/Nissan) salesman.
5A. The inspector discovered a problem in the mansion's (bathroom/basement/kitchen).
5B. The old mansion had a major (leak/pest/mold) problem.
6A. A front-page photograph of a (accident/burglary/wildfire) won an award.
6B. The (Journal/Gazette/Observer)'s photography won an award.
7A. Andrea had to suffer through a (rerouted/canceled/delayed) flight while traveling this month.
7B. Andrea encountered flight problems on her way to a (meeting/wedding/funeral).
8A. The leader of the rock band changed roles when someone else took over on (vocals/guitar/piano).
8B. After joining the rock band, (Laura/Chris/David) took over one of the bandleader's roles.
9A. The forthcoming film ran into trouble when the (director/producer/star) quit.
9B. Someone quit the forthcoming film because of a dispute over the (casting/effects/screenplay).
10A. The (teachers/students/janitors) at the school were particularly pleased by the renovations.
10B. People at the school were particularly pleased by the (warmer/larger/brighter) rooms.
11A. The (bride/groom/usher) had a connection that helped the couple reserve a place for their wedding.
11B. The couple reserved a (resort/chapel/park) for their wedding on Valentinfe's Day.
12A. The ecologists discovered the (rhino/hippo/elephant) herds were getting larger.
12B. The ecologists discovered the animal populations were getting larger in (Kenya/Sudan/the Congo).
13A. Jennifer's (cat/dog/mouse) had a problem with one of her cousin's pets.
13B. Chaos broke out at Jennifer's house because of her cousin's (ferret/rabbit/hamster).
14A. Critics were especially pleased by the new book about the (Depression/Revolution/Civil War).
14B. The history book that the critics liked best was from the publisher in (Seattle/Houston/Boston).
15A. To get more exercise, Mike decided to (bike/walk/jog) to work.
15B. To get more exercise, Mike decided to (swim/hike/climb) on the weekends.
16A. At the high school reunion, the friends were amused by (Jessica/Lindsay/Michelle)'s career choice.
16B. At the high school reunion, the friends were surprised that one of them was now (a lawyer/a banker/an accountant).
17A. The (French/German/English) knight married the baroness's daughter.
17B. The competition to marry the baroness's daughter was resolved by the (archery/fencing/jousting) contest.
18A. Samantha went to a study group for her (biology/physics/chemistry) midterm.
18B. Samantha's study group met at the (cafe/quad/library).
19A. Tina and her boyfriend made their decision about what to do on Friday by choosing the (shorter/cheaper/closer) event.
19B. Tina and her boyfriend didn't decide to attend the (film/play/concert) until Friday morning.
20A. The (professors/architects/electricians)' bowling league was temporarily banned from the local bowling lanes due to a rule violation.
20B. The management of the bowling lanes had to ban one of the leagues for bringing in (snacks/drinks/cigars).
21A. The only bad weather during Julia's trip to the mountains was the (rain/snow/hail).
21B. The only bad weather during Julia's trip to the mountains was on (Monday/Sunday/Saturday).
22A. On the game show, Ben chose the prize behind the (black/white/brown) door.
22B. The door that Ben opened on the game show had a (goat/boat/coat) behind it as the prize.
23A. NASA lost some of the (videos/photos/measurements) from the space probe due to a bug.
23B. NASA lost some of the data from (Neptune/Jupiter/Saturn) due to a bug in the space probe.
24A. The quarterback blamed a blown (passing/rushing/kicking) play on the fact that he hadn't polished everything for good luck.
24B. The quarterback blamed the team's performance during the second quarter on the fact that he hadn't polished his (shoes/rings/helmet) before the game.
25A. The fashion designer's (shirts/skirts/shorts) were a hit at the show.
25B. The fashion designer unveiled his new line at a show in (Paris/Milan/London).
26A. The sporting goods manufacturer decided to sign the (pitcher/swimmer/golfer) to endorse one of its products.

26B. The athlete endorsing the (watches/jackets/drinks) was dropped after a scandal.
27A. Elizabeth chose a (spice/lemon/cherry) cake for her friend's birthday party.
27B. Elizabeth served (cider/punch/tea) for her friend's birthday party.
28A. The easiest way for Bridget and her niece to reach the museum was by (bus/car/train).
28B. Bridget took her niece to the (science/history/art) museum.
29A. Suzanne's first job as a translator at the McKee Company was to translate the (invoice/letters/contract).
29B. The McKee Company hired Suzanne to translate documents from (Russian/Chinese/Korean).
30A. The haunted house fundraiser was organized by the (seniors/juniors/sophomores).
30B. The haunted house fundraiser raised money to fight (cancer/hunger/crime).
31A. The bird that Jason spotted was one of the (ducks/swans/loons).
31B. Jason spotted a bird at the (river/marsh/lake).
32A. The (American/Japanese/Canadian) engineers designed the most successful computer chip.
32B. The computer chip that was most successful was the (smallest/quickest/coolest) of any available.
33A. Because the critic caught the flu, he only had a chance to visit the (Mexican/Italian/Indian) restaurant.
33B. The food critic gave a favorable review to the (desserts/specials/entrees) at one of the new restaurants.
34A. The winner of the Pulitzer Prize for (journalism/literature/music) declined to attend one of the celebratory events.
34B. The (dinner/parade/press conference) was not attended by one of the city's Pulitzer Prize winners.
35A. In the new children's movie, the crops are struck by (drought/disease/a tornado).
35B. In the new children's movie, a talking (cow/pig/horse) saves the farm.

Appendix C

C.1. Stimuli for Experiment 3

1. *Context:* The old mansion was finally sold to a new owner, but was in a state of disrepair. The handyman had fixed the (leak/mold) problem in the (kitchen/basement), but the new owner suspected there might be other problems too. An inspector was sent in to check for pests and (mold/leak)s in both the bathroom and (basement/kitchen).
Continuation: She discovered a major pest problem in the mansion's bathroom and said a specialist would be needed to fix it. It was one of the worst messes she'd ever seen.
2. *Context:* After Steve and his wife took a trip to the (Redwoods/Cascades) last (autumn/summer), they decided that they wanted to visit the Badlands and the (Cascades/Redwoods), too. Steve could take his vacation in either the (summer/autumn) or the spring.
Continuation: Steve and his family eventually decided to travel to the Badlands in the spring because they found a good deal on a hotel.
3. *Context:* Annette and her (mother/sister) completed all their holiday shopping for Annette's (sister/mother) and niece while on vacation in Spain. Annette stopped in a gift shop in Barcelona and bought a mug and a (purse/book), as well as a (book/purse) for herself.
*Continuation:*She gave her niece the mug for Christmas. She was very happy with the gift and said it was her favorite of the gifts she received that year.
4. *Context:* The local parks commission had a busy meeting on Wednesday to decide how to spend its money for the year. Some money had been set aside to (replant/repair) the parks and build a (skatepark/fountain), but people disagreed on whether the remaining budget should be used to (repair/replant) or expand the parks. The mayor wanted to add a (fountain/skatepark) but City Council wanted a playground.
Continuation: After a long debate, a compromise was made to expand the parks and build a playground but it didn't seem like anyone was very happy about the decision.
5. *Context:* Two best friends from college, Matt and (Eric/Nick), ended up as rival salesmen selling Toyota and (Subaru/Nissan) cars. They had a friendly competition going and their friend (Nick/Eric) at the (Nissan/Subaru) dealership even made a silly trophy for it.
Continuation: Most years, the trophy went to Matt because of his Toyota deals.
6. *Context:* Andrea hated flying ever since her flight was (rerouted/delayed) when she was trying to get to a (funeral/wedding). But with both a (wedding/funeral) and a conference to attend this month, she knew she'd just have to endure it. She only hoped her flights weren't canceled or (delayed/rerouted).
Continuation: Unfortunately, her fears came true when the airline canceled her flight to the conference with very little notice. Andrea wrote an angry letter to the airline about the situation.
7. *Context:* When the local rock band first formed in (Chicago/Normal), the bandleader only handled (drums/guitar) and recruited members from (Normal/Chicago) and Urbana to handle vocals

(*continued on next page*)

and (guitar/drums).
Continuation: But when one of them graduated and moved back to Urbana for a job, the bandleader took over on vocals as well.

8. *Context:* The (director/producer) and star of a forthcoming film both liked the (effects/casting) but were having a big dispute over the (casting/effects) and screenplay. Although the (producer/director) tried to mediate the argument,
Continuation: they failed to reach consensus on the screenplay and the star quit in a rage.
9. *Context:* The (janitors/students) were unhappy with the renovation of the school, because they had hoped for (brighter/warmer) rooms. But both the teachers and the (students/janitors) were rather happy with the fact that the rooms were now larger and (warmer/brighter) for the dreary winter months.
Continuation: The teachers were particularly pleased to have the larger rooms, which had been their top priority.
10. *Context:* Some ecologists from (Sudan/Gabon) went to check on the rhino and (hippo/elephant) populations in (Gabon/Sudan) and Kenya. They got to see a baby (elephant/hippo) and
Continuation: were happy to discover that the rhino herds in Kenya were getting larger.
11. *Context:* Dorothy and her friends formed a book club where they could read mystery and (fantasy/romance) novels without having to endure any (romance/fantasy) books. Dorothy agreed to host the meetings at her home, but asked her friends to bring (cheese/shrimp), nuts, and other snacks–but no (shrimp/cheese); she was allergic.
Continuation: Their first meeting got off to a slow start because some people were late, but Dorothy nibbled on some nuts while waiting to discuss the mystery they'd read.
12. *Context:* Gyro-Tek's new product line had done poorly in test marketing. The CEO blamed the (production/marketing) team for the problems and immediately ruled out their suggestion to (delay/cancel) the products. Instead, he met with the (marketing/production) and research teams to get their input on whether to (cancel/delay) or revise the product roll-out.
Continuation: Eventually, he decided to heed the advice of the research team and revise the product line.
13. *Context:* Many of Channel 3's news staff were already covering a big (accident/wildfire). So when word about the burglary and the (wildfire/accident) reached the station, there was no (editor/cameraman) available. But the manager assigned the remaining reporter and (cameraman/editor) to get started on one of the new stories anyway.
Continuation: This turned out to be a good decision, because the reporter's work on the burglary story later won an award.
14. *Context:* Jennifer owned both a cat and a (dog/rat), but the two pets got along great with each other. They each had their own (ball/sock) and chew toys and neither of them was interested in the one (sock/ball) toy. But whenever Jennifer's boyfriend brought his (rat/dog) over to her apartment,
Continuation: it got into fights with Jennifer's cat over the chew toy, and chaos broke out.
15. *Context:* Strangely enough, on the same day, both the Seattle publisher and the (Boston/Houston) publisher released new biographies of Presidents (Lincoln/Clinton) and Kennedy. To get an opinion on the books, (Houston/Boston) radio interviewed a local professor who was better known for her research on President (Clinton/Lincoln).
Continuation: She judged the book on Kennedy from the Seattle publisher to be the best of the bunch.
16. *Context:* Mike's doctor told him that he should get more exercise and recommended (hiking/cycling) and a (swim/health) club. Mike couldn't see himself doing either of those. But he was willing to consider (cycling/hiking) or jogging, and maybe a (health/swim) or tennis club.
Continuation: First he started jogging on the weekends, and then signed up for a trial tennis club membership. When he went back in a few months, the doctor seemed to think the plan was working for Mike.
17. *Context:* A lot of people at the high school reunion were surprised by their classmates' career choices. (Maggie/Rachel) had become a successful (banker/farmer). And while Ashley and (Rachel/Maggie) had both claimed they'd never want to be a lawyer or a (farmer/banker),
Continuation: sure enough, Ashley was now a lawyer in Wisconsin. They all had a good laugh about it.
18. *Context:* To win the hand of the (English/French) baron's daughter, the German and the (French/English) knights competed in a tournament of (archery/jousting) and fencing. Both knights gave the two events their best effort even though they were really better at (jousting/archery).
Continuation: Eventually, the German knight emerged victorious by winning the fencing event and married the daughter.
19. *Context:* The local bowling lanes had gotten to be quite popular. The owner, a retired (architect/carpenter), kept the lanes inexpensive and allowed people to bring in their (phones/drinks). The only rules were that no snacks or (drinks/phones) were allowed. Thanks to these friendly policies, bowling leagues had been formed by both the (carpenter/architect)s and the

professors.
Continuation: But when the professors kept bringing in their snacks every week, management had no choice but to temporarily ban them from the lanes.

20. *Context:* Molly was on a game show and had to pick one of three boxes to get a prize. The host gave her a hint by telling her that inside the (brown/green) box was just a cheap, old (microwave/necklace). That left the (green/brown) box and the white box. One of them had a (necklace/microwave) and one had a computer.
Continuation: Nervously, Molly opened the white box and discovered that the computer was inside. The audience went wild.
21. *Context:* Originally, the space probe Cosmo III was designed to fly past Jupiter and (Saturn/Neptune) and send photos and (videos/measurements) back to NASA from both planets. NASA needed this information to guide the (measurements/videos) they were going to take of (Neptune/Saturn) on a future mission.
Continuation: However, due to a glitch in the programming of the Cosmo III, it lost the photos taken of Jupiter and put the future mission in trouble.
22. *Context:* The McKinley High Bruins football team had a lot of superstitions. For good luck, one receiver always tapped his (shoes/helmet) against his locker. And, the quarterback thought that if he didn't rub his (helmet/shoes) and rings, the team's rushing and (kicking/passing) wouldn't go as well. On the strength of their (passing/kicking) game, the team took the lead early this Friday.
Continuation: But when the team blew a big rushing play during the second quarter, the quarterback blamed it on the fact that he had forgotten to rub his rings the night before.
23. *Context:* The renowned fashion designer from (Paris/London) saw his line of (skirts/shorts) fail to sell. Hoping to redeem his reputation, he divided his time between Milan and (London/Paris) while working on lines of (shorts/skirts) and shirts.
Continuation: His shirts were a hit when he revealed them at a Milan fashion show, but the others were less favorably received.
24. *Context:* The president of Acme Sporting Goods, who was a former (pitcher/golfer), wanted the company to expand beyond its lines of (watches/drinks). He looked for some athlete endorsements that they could use to promote the (drinks/watches) and jackets that they planned to introduce. He met with a (golfer/pitcher) and a swimmer.
Continuation: The company signed the swimmer to endorse the jackets in a commercial, but they dropped him after he became involved in a scandal.
25. *Context:* Elizabeth was in charge of organizing her friend Mary's birthday party. Last year, they'd had (cherry/spice) cake and (punch/cider) and Mary had hated it. So, Elizabeth asked Mary's brother if she should serve tea or (cider/punch) instead and if Mary would prefer lemon cake or (spice/cherry) cake.
Continuation: Mary's brother suggested tea but didn't know about the cake. Elizabeth picked out a lemon cake, which she personally liked best.
26. *Context:* Bridget's granddaughter was getting in on the (bus/train) this evening for a visit. Bridget knew her granddaughter hated (science/history) but might be interested in the art and (history/science) museums in town. She wasn't sure about the traffic, so she checked online to see if it would be easier to go by (train/bus) or by car.
Continuation: They could easily reach the art museum by car right as it opened.
27. *Context:* The McKee Company was trying to expand into the Chinese and (Korean/Russian) markets, but confusion arose in the first (invoices/contracts) they exchanged with their partners. They realized they needed a better translator. They found their ideal candidate in the daughter of a (Russian/Korean) diplomat who had lived all over the world. When the company was ready to send out new (contracts/invoices) and letters,
Continuation: her first job was to make sure the translation of the letters into Chinese was clear.
28. *Context:* Every year on Service Day, the campus held a charity fun run. Last year, the (faculty/seniors) raised a lot of money for the fight against (hunger/crime). This year, they weren't participating, so it was up to the (seniors/faculty) and juniors to raise money for victims of (crime/hunger) and cancer.
Continuation: The juniors won because they were campaigning against cancer and that attracted a lot of donations.
29. *Context:* A (Japanese/Canadian) phone company was reviewing bids from groups of American and (Canadian/Japanese) engineers to produce computer chips for its new phones. The engineers were asked to make the chips smaller and (cool/fast)er while still keeping them (fast/cool).
Continuation: The design by the American engineers made the most improvement in making the chip smaller and so they got the job.
30. *Context:* A new Mexican and a new (Indian/Italian) restaurant had recently opened. Both were waiting to hear whether or not the local food critic would like their (desserts/entrees) and specials. They were nervous because the critic was notoriously harsh and disliked even the

(*continued on next page*)

popular (entrees/desserts) at the local (Italian/Indian) restaurant. The critic originally planned to dine at both new restaurants during the week. *Continuation:* But he caught a cold and could only visit the Mexican restaurant, where he awarded the specials a favorable review.

31. *Context:* Samantha's (biology/physics) and chemistry midterms were both the same week. She was stressed out, especially after getting a C on her (physics/biology) midterm last week. She tried both the cafe and the (quad/library) to study but couldn't concentrate. But, things started going better when she ran into three of her classmates at the (library/quad). *Continuation:* They invited her to a chemistry study group at the cafe and she felt more confident.
32. *Context:* A new children's movie has just been released about Happy (Hen/Cow) Farm, where the (cow/hen)s and pigs can all talk. At the start of the movie, the farm survives a (tornado/disease). In the aftermath, the farmers worry that the crops could be finished off by a (disease/tornado) or a drought. *Continuation:* But when a severe drought strikes, a heroic pig ends up saving the farm from bankruptcy.
33. *Context:* Brad hadn't been finishing in a long time, but when he saw the nice (perch/trout) that his friend Duane reeled in yesterday (afternoon/evening), he decided to get out his fishing pole again. He spent all morning and (evening/afternoon) in his boat, hoping to catch some bass or (trout/perch). *Continuation:* He didn't have the best haul, but the bass that he reeled in during the morning was the biggest he'd ever caught.
34. *Context:* Jason's girlfriend invited him to go birdwatching. He agreed because he remembered how excited she was to see some (swans/loons) while hiking near the (marsh/lake) recently. Jason was expecting it would be easy to spot all the ducks and (loons/swans) down at the (lake/marsh) and river. But it was so early in the morning that he was too sleepy to do so. *Continuation:* When he beat the rest of the group by spotting one of the ducks at the river first, he was excited and told his girlfriend he'd be willing to go birdwatching again.
35. *Context:* It was (hail/snow)ing hard. Julia checked the weather forecast because she had to drive to a job interview in the morning. The forecast was hard (snow/hail) or rain tomorrow. Julia was worried because it was a long drive to the interview and her (bumper/airbag) was broken. At least she had her seatbelt and (airbag/bumper) still. *Continuation:* Julia made it to the job interview, but her car slid on the drive home through the rain and hit a tree. Luckily, Julia was protected by her seatbelt and she was just fine.
36. *Context:* Erin expected a boring train ride home for the weekend because she'd forgotten to bring the (newspaper/magazine) and (chips/fruit) that she usually liked to take along. She bought some (fruit/chips) and candy at the station and got out the (magazine/newspaper) and homework that she did remember to take. *Continuation:* She occupied herself with the homework and snacked on candy until she got tired, and then phoned a friend.

Appendix D

D.1. Test probes for Experiment 3

1A. A specialist was needed to fix the old mansion's (pest/mold/leak) problem.
1B. The inspector discovered a problem in the mansion's (bathroom/basement/kitchen) that needed a specialist.
2A. Steve and his family decided the next park they would visit would be the (Badlands/Cascades/Redwoods).
2B. Steve and his family decided to visit another national park during the (spring/summer/autumn).
3A. Annette gave her (niece/sister/mother) a gift from the gift shop in Barcelona that she was very happy to get.
3B. Annette's gift of a (mug/purse/book) from Barcelona was a favorite for its recipient.
4A. After a debate, the parks commission decided to use its remaining budget to (expand/repair/replant) the parkland.
4B. After a debate, the parks commission decided to use its remaining budget build a (playground/fountain/skatepark).
5A. Most years, the winner of the sales contest between college friends was (Matt/Eric/Nick).
5B. Most years, the winner of the sales contest between college friends was the (Toyota/Subaru/Nissan) salesman.
6A. Andrea had to suffer through a (canceled/delayed/rerouted) flight while traveling this month.
6B. Andrea encountered a sudden change of travel plans on her way to a (conference/wedding/funeral).
7A. One of the members of the rock band graduated and moved back to (Urbana/Normal/Chicago).
7B. The leader of the rock band took over on (vocals/guitar/drums) after someone moved away.
8A. Someone quit the forthcoming film because of a dispute over the (screenplay/casting/effects).

8B. The forthcoming film ran into trouble when the (star/director/producer) quit.
9A. The (teachers/students/janitors) at the school were particularly pleased by the renovations.
9B. People at the school were particularly pleased by the (larger/warmer/brighter) rooms.
10A. The ecologists discovered the (rhino/hippo/elephant) herds were getting larger.
10B. The ecologists discovered the animal populations were getting larger in (Kenya/Gabon/Sudan).
11A. The book club members enjoyed (nuts/cheese/shrimp) at their first meeting.
11B. The first book that the book club read was a (mystery/fantasy/romance).
12A. Gyro-Tek's CEO made his decision about the product line based on the advice of the (research/marketing/production) team.
12B. After the new product line tested poorly, Gyro-Tek's CEO decided to (revise/cancel/delay) it.
13A. A (reporter/cameraman/editor) at Channel 3 won an award.
13B. Channel 3's work on the (burglary/wildfire/accident) story won an award.
14A. Jennifer's (cat/dog/rat) would get into fights with her boyfriend's pet.
14B. Jennifer's pet fought her boyfriend's pet over the (chew/ball/sock) toy.
15A. The history professor especially liked the new biography of President (Kennedy/Lincoln/Clinton).
15B. The history book that the professor liked best was from the publisher in (Seattle/Boston/Houston).
16A. To get more exercise, Mike tried (jogging/cycling/hiking) on the weekends.
16B. To get more exercise, Mike joined a (tennis/health/swim) club.
17A. At the high school reunion, the friends were amused by (Ashley/Rachel/Maggie)'s career choice.
17B. At the high school reunion, the friends were amused that one of them was now a (lawyer/farmer/banker).
18A. The knight who married the baron's daughter was (German/French/English).
18B. The competition to marry the baron's daughter was resolved by the (fencing/archery/jousting) contest.
19A. The (professors/carpenters/architects)' bowling league was temporarily banned due to a rule violation.
19B. One of the bowling leagues was banned for bringing in (snacks/drinks/phones).
20A. On the game show, Molly chose the prize in the (white/green/brown) box.
20B. The box that Molly opened on the game show had a new (computer/necklace/microwave) inside.
21A. NASA lost some of the (photos/videos/measurements) from the space probe due to a bug.
21B. NASA lost some of the data from (Jupiter/Saturn/Neptune) due to a bug in the space probe.
22A. The McKinley High Bruins blew a big (rushing/kicking/passing) play.
22B. The quarterback blamed the team's mistakes on the fact that he hadn't rubbed his (rings/helmet/shoes) the night before the game.
23A. The fashion designer's (shirts/shorts/skirts) were a hit at the show.
23B. The fashion designer unveiled his new line at a show in (Milan/London/Paris).
24A. The sporting goods manufacturer signed a (swimmer/golfer/pitcher) to endorse some of its products.
24B. The athlete endorsing the (jackets/drinks/watches) was dropped after a scandal.
25A. Elizabeth served (tea/cider/punch) for her friend's birthday party.
25B. Elizabeth chose a (lemon/spice/cherry) cake for her friend's birthday party.
26A. Bridget took her granddaughter to the (art/history/science) museum.
26B. The easiest way for Bridget and her granddaughter to reach the museum was by (car/train/bus).

(*continued on next page*)

Table E1
Summary of random item and participant effects and correlations in model of "True" responses in Experiment 1.

Random effect	s^2	Correlations				
		1	2	3	4	5
Item						
1. Intercept	0.18	–				
2. Emphasized word	0.10	−.32	–			
3. Rejections of unmentioned probes	2.76	.11	−.30	–		
4. Rejections of alternative probes	5.94	−.03	<.01	.41	–	
5. Emphasis × rejections of unmentioned probes	6.20	.47	−.79	.67	.47	–
6. Emphasis × rejections of alternative probes	5.43	−.40	−.37	.03	−.59	.02
Participant						
1. Intercept	0	.01				

Table E2
Summary of random item and participant effects and correlations in model of log reading time in Experiment 2.

Random effect	s^2	Correlations		
		1	2	3
Item				
1. Non-emphasized critical word (baseline)	0.005	–		
2. Emphasized word	0.002	.29	–	
3. Spillover region	0.004	−.43	−.18	–
4. Spillover region × emphasized word	0.003	−.73	−.76	>−.01
Participant				
1. Non-emphasized critical word (baseline)	0.081	–		
2. Emphasized word	0.001	.68	–	
3. Spillover region	<0.001	<−.99	−.67	–
4. Spillover region × emphasized word	0.007	−.68	−.91	.67
Residual variance	0.121			

Table E3
Summary of random item and participant effects and correlations in models of "True" responses in Experiment 2.

Random effect	s^2	Correlations	
		1	2
Item			
1. Intercept	0.29	–	
2. Rejections of unmentioned probes	1.71	−.91	–
3. Rejections of alternative probes	5.52	.14	.06
Participant			
1. Intercept	0.31	–	
2. Rejections of unmentioned probes	0.83	−.69	–
3. Rejections of alternative probes	0.03	−.78	.99

Table E5
Summary of random item and participant effects and correlations in models of "True" responses in Experiment 3.

Random effect	s^2	Correlations	
		1	2
Item			
1. Intercept	0.35	–	
2. Rejections of merely mentioned probes	1.25	−.42	–
3. Rejections of alternative probes	1.20	−.31	−.07
Participant			
1. Intercept	0.07	–	
2. Rejections of merely mentioned probes	0.60	−.96	–
3. Rejections of alternative probes	0.33	−.57	.78

27A. The new translator's first job at the McKee Company was to translate the (letters/contracts/invoices).
27B. At the McKee Company, the new hire's first job was to translate some documents into (Chinese/Korean/Russian).
28A. This year's charity fun run was won by the (juniors/seniors/faculty).
28B. The winning team in the charity fun run attracted a lot of donations to fight (cancer/crime/hunger).
29A. The job to design the computer chips was awarded to (American/Canadian/Japanese) engineers.
29B. The winning computer chip design made the most improvement in making the chip (smaller/cooler/faster).
30A. Because the food critic caught a cold, he only

Table E4
Summary of random item and participant effects and correlations in model of log reading time in Experiment 3.

Random effect	s^2	Correlations		
		1	2	3
Item				
1. Non-emphasized critical word (baseline)	0.002	–		
2. Emphasized word	0.011	−.24	–	
3. Spillover region	0.005	.50	−.14	–
4. Spillover region × emphasized word	0.008	.21	.96	.36
Participant				
1. Non-emphasized critical word (baseline)	0.094	–		
2. Emphasized word	0.021	.72	–	
3. Spillover region	0.001	−.78	−.64	–
4. Spillover region × emphasized word	0.026	−.81	−.83	.96
Residual variance	0.140			

visited the (Mexican/Indian/Italian) restaurant.
30B. The food critic gave a favorable review to the (specials/desserts/entrees) at one of the new restaurants.
31A. Samantha went to a study group for her (chemistry/biology/physics) midterm.
31B. Samantha's midterm study group met at the (cafe/quad/library).
32A. In the new children's movie, a talking animal saves the farm from a (drought/disease/tornado).
32B. In the new children's movie, a talking (pig/cow/hen) saves the farm.
33A. Brad caught a huge (bass/trout/perch) at the lake this past weekend.
33B. When Brad went fishing at the lake, he caught a huge fish during the (morning/evening/afternoon).
34A. The bird that Jason spotted first was one of the (ducks/loons/swans).
34B. Jason spotted a bird at the (river/lake/marsh).
35A. Julia's car slid while driving through (rain/snow/hail).
35B. Julia was protected in her car accident by her (seatbelt/airbag/bumper).
36A. On her train ride home, Erin kept herself busy with her (homework/magazine/newspaper).
36B. On her train ride home, Erin snacked on (candy/fruit/chips).

Appendix E

Tables E1–E5.

References

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Almor, A., & Eimas, P. D. (2008). Focus and noun phrase anaphors in spoken language comprehension. *Language and Cognitive Processes, 23*, 201–225. http://dx.doi.org/10.1080/01690960701330936.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412. http://dx.doi.org/10.1016/jmla.2007.12.005.

Bates, D., Maechler, M., & Bolker, B. (2011). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.99375-39. Retrieved from http://CRAN.R-project.org/package-lme4.

Beckman, M. E., & Elam, G. A. (1997). *Guidelines for ToBI labelling (version 3.0)*. Retrieved from http://www.ling.ohio-state.edu/~tobi/ame_tobi/labelling_guide_v3.pdf.

Benjamin, A. S. (2005). Response speeding mediates the contribution of cue familiarity and target retrievability to metamnemonic judgments. *Psychonomic Bulletin & Review, 12*, 874–879. http://dx.doi.org/10.3758/BF03196779.

Benjamin, A. S. (2008). Memory is more than just remembering: Strategic control of encoding accessing memory and making decisions. In A. S. Benjamin & B. H. Ross (Eds.). *The psychology of learning and motivation: Skill and strategy in memory use* (Vol. 48, pp. 175–223). London: Academic Press. doi:10.106/S0079-7421(07)48005-7.

Besner, D. (1987). Phonology, lexical access, and articulatory suppression: A critical review. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 39*, 467–478. http://dx.doi.org/10.1080/14640748708401799.

Birch, S., & Rayner, K. (2010). Effects of syntactic prominence on eye movements during reading. *Memory & Cognition, 38*, 740–752. http://dx.doi.org/10.3758/MC.38.6.740.

Blok, P. I., & Eberle, K. (1999). What is the alternative? The computation of focus alternatives from lexical and sortal information. In P. Bosch & R. van der Sandt (Eds.), *Focus: Linguistic, cognitive, and computational perspectives* (pp. 105–119). Cambridge, United Kingdom: Cambridge University Press.

Bock, J. K., & Mazzella, J. E. (1983). Intonational marking of given and new information: Some consequences for comprehension. *Memory & Cognition, 11*, 64–76. http://dx.doi.org/10.3758/BF03197663.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436. http://dx.doi.org/10.1163/156856897X00357.

Brainerd, C. J., Reyna, V. F., & Estrada, S. (2006). Recollection rejection of false narrative statements. *Memory, 14*, 672–691. http://dx.doi.org/10.1080/09658210600648449.

Breen, M., & Clifton, C. Jr., (2010). Stress matters: Effects of anticipated lexical stress on silent reading. *Journal of Memory and Language, 64*, 153–170. 10.106/jml.2010.11.001.

Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science, 32*, 643–684. http://dx.doi.org/10.1080/03640210802066816.

Byram Washburn, M., Kaiser, E., & Zubizarreta, M. L. (2011). Focus facilitation and non-associative sets. In R. Artstein, M. Core, D. DeVault, K. Georgila, E. Kaiser, & A. Stent (Eds.), SemDial 2011: Proceedings of the 15th workshop on the semantics and pragmatics of dialogue (pp. 94–102).

Calhoun, S. (2009). What makes a word contrastive? Prosodic, semantic, and pragmatic perspectives. In D. Barth-Weingarten, N. Dehé, & A. Wichmann (Eds.), *Studies in pragmatics: Where prosody meets pragmatics: Research at the interface* (pp. 53–77). Bingley, UK: Emerald.

Caplan, D., DeDe, G., Waters, G., Michaud, J., & Tripodis, Y. (2011). Effects of age, speed of processing, and working memory on comprehension of sentences with relative clauses. *Psychology and Aging, 26*, 439–450. http://dx.doi.org/10.1037//a0021837.

Christianson, K., & Luke, S. G. (2011). Context strengthens initial misinterpretations of text. *Scientific Studies of Reading, 15*, 136–166. http://dx.doi.org/10.1080/10888431003636787.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335–359. http://dx.doi.org/10.1016/S0022-5371(73)80014-3.

Clifton, C., Jr., Bock, J., & Radó, J. (2000). Effects of the focus particle *only* and intrinsic contrast on comprehension of reduced relative clauses. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 591–619). Amsterdam: Elsevier.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7*, 249–253. http://dx.doi.org/10.1177/014662168300700301.

Daneman, M., Lennertz, T., & Hannon, B. (2007). Shallow semantic processing of text: Evidence from eye movements. *Language and Cognitive Processes, 22*, 83–105. http://dx.doi.org/10.1080/01690960500372725.

Davis, T., & Love, B. C. (2010). Memory for category information is idealized through contrast with competing options. *Psychological Science, 21*, 234–242. http://dx.doi.org/10.1177/0956797609357712.

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition, 118*, 111–115. http://dx.doi.org/10.1016/j.cognition.2010.09.012.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24*, 523–533. http://dx.doi.org/10.3758/BF03200940.

Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition, 25*, 691–700. http://dx.doi.org/10.3758/BF03211311.

Ferreira, F., & Clifton, C. Jr., (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348–368. http://dx.doi.org/10.1016/0749-596X(86)90006-9.

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass, 1*, 71–83. http://dx.doi.org/10.1111/j.1749-818x.2007.00007.x.

Filik, R., Purdy, K., Gale, A., & Gerrett, D. (2006). Labeling of medicines and patient safety: Evaluating methods of reducing drug name confusion. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 48*, 39–47. http://dx.doi.org/10.1518/001872006776412199.

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research, 27*, 285–319. http://dx.doi.org/10.1023/A:1023258301588.

Font (2011). *Oxford English Dictionary* http://www.oed.com.proxy2.library.illinois.edu/view/Entry/73956 Retrieved 11.08.2011.

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language, 63*, 367–386. http://dx.doi.org/10.1016/j.jml.2010.06.004.

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2012). The effects of age on the strategic use of pitch accents in memory for discourse: A processing-resource account. *Psychology and Aging, 27*, 88–98. http://dx.doi.org/10.1037/a0024138.

Freeman, E., Heathcote, A., Chalmers, K., & Hockley, W. (2010). Item effects in recognition memory for words. *Journal of Memory and Language, 62*, 1–18. http://dx.doi.org/10.1016/j.jml.2009.09.004.

Gao, X., Stine-Morrow, E. A. L., Noh, S. R., & Eskew, R. T. Jr., (2011). Visual noise disrupts conceptual integration in reading. *Psychonomic Bulletin & Review, 18*, 83–88. http://dx.doi.org/10.3758/s13423-010-0014-4.

Golding, J. M., & Fowler, S. B. (1992). The limited facilitative effect of typographical signals. *Contemporary Educational Psychology, 17*, 99–113. http://dx.doi.org/10.1016/0361-476X(92)90052-Z.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Gundel, J. K. (1999). On different kinds of focus. In P. Bosch & R. van der Sandt (Eds.), *Focus: Linguistic, cognitive, and computational perspectives* (pp. 293–305). Cambridge, United Kingdom: Cambridge University Press.

Gussenhoven, C. (1983). *A semantic analysis of the nuclear tones of English (Doctoral dissertation)*. Bloomington, IL: Indiana University Linguistics Club.

Halliday, M. A. K. (1967). Notes on transitivity and theme in English: Part 2. *Journal of Linguistics, 3*, 177–274. http://dx.doi.org/10.1017/S0022226700016613.

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90*, 414–434. http://dx.doi.org/10.1037/0022-0663.90.3.414.

Hartley, J., Bartlett, S., & Branthwaite, A. (1980). Underlining can make a difference—sometimes. *The Journal of Educational Research, 73*, 218–224.

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 13*, 512–521. http://dx.doi.org/10.1016/S0022-5371(74)80003-4.

Henderson, J. M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 417–429. http://dx.doi.org/10.1037/0278-7393.16.3.417.

Hunt, R. R., & Lamb, C. A. (2001). What causes the isolation effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1359–1366. http://dx.doi.org/10.1037/0278- 7393.27.6.1359.

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language, 58*, 541–573. http://dx.doi.org/10.1016/j.jml.2007.06.013.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformations or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446. http://dx.doi.org/10.1016/j.jml.2007.11.007.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*, 228–238. http://dx.doi.org/10.1037/0096-3445.111.2.228.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14*, 219–224. http://dx.doi.org/10.3758/BF03194055.

Ledoux, K., Gordon, P. C., Camblin, C. C., & Swaab, T. Y. (2007). Coreference and lexical repetition: Mechanisms of discourse integration. *Memory & Cognition, 35*, 801–815. http://dx.doi.org/10.3758/BF03193316.

Lorch, R. F., Jr., Lorch, E. P., & Klusewitz, M. A. (1995). Effects of typographical cues on reading and recall of text. *Contemporary Educational Psychology, 20*, 51–64. http://dx.doi.org/10.1006/ceps.1995.1003.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory* (2nd ed.). New York: Erlbaum.

Matzen, L. E., & Benjamin, A. S. (2009). Remembering words not presented in sentences: How study context changes patterns of false memories. *Memory & Cognition, 37*, 52–64. http://dx.doi.org/10.3758/MC.37.1.52.

Matzen, L. E., Taylor, E. G., & Benjamin, A. S. (2011). Contributions of familiarity and recollection rejection to recognition: Evidence from the time course of false recognition for semantic and conjunction lures. *Memory, 19*, 1–16. http://dx.doi.org/10.1080/09658211.2010.530271.

McAteer, E. (1992). Typeface emphasis and information focus in written language. *Applied Cognitive Psychology, 6*, 345–359. http://dx.doi.org/10.1002/acp.2350060406.

Nieuwland, M. S., Otten, M., & Van Berkum, J. J. A. (2007). Who are you talking about? Tracking discourse-level referential processing with event-related brain potentials. *Journal of Cognitive Neuroscience, 19*, 228–236. http://dx.doi.org/10.1162/jocn.2007.19.2.228.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442. http://dx.doi.org/10.1163/156856897X00357.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intention in communication* (pp. 271–311). Cambridge, MA: MIT Press.

Quené, H., & van den Bergh, H. (2004). On multi-level modeling of data from repeated measures design: a tutorial. *Speech Communication, 43*, 103–121. http://dx.doi.org/10.1016/j.specom.2004.02.004.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422. http://dx.doi.org/10.1037/0033-2909.124.3.272.

Reder, L. M., & Kusbit, G. W. (1991). Locus of the Moses illusion: Imperfect encoding, retrieval, or match? *Journal of Memory and Language, 30*, 385–406. http://dx.doi.org/10.1016/0749- 596X(91)90013-A.

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review, 16*, 1–21. http://dx.doi.org/10.3758/PBR.16.1.1.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics, 1*, 75–116. http://dx.doi.org/10.1007/BF02342617.

Sanford, A. J. S., Price, J., & Sanford, A. J. (2009). Enhancement and suppression effects resulting from information structuring in sentences. *Memory & Cognition, 37*, 880–888. http://dx.doi.org/10.3758/MC.37.6.880.

Sanford, A. J. S., Sanford, A. J., Molle, J., & Emmott, C. (2006). Shallow processing and attention capture in written and spoken discourse. *Discourse Processes, 42*, 109–130. http://dx.doi.org/10.1207/s15326950dp4202_2.

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: not noticing the evidence. *TRENDS in Cognitive Sciences, 6*, 382–386. http://dx.doi.org/10.1016/S1364-6613(02)01958-7.

Sedivy, J. C. (2002). Invoking discourse-based contrast sets and resolving syntactic ambiguities. *Journal of Memory and Language, 46*, 341–370. http://dx.doi.org/10.1006/jmla.2001.2812.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71*, 109–147. http://dx.doi.org/10.1016/S0010-0277(99)00025-6.

Selkirk, E. (1995). Sentence prosody: Intonation, stress, and phrasing. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (Vol. 1, pp. 550–569).

Selkirk, E. (2002). Contrastive *FOCUS* vs. presentational *focus*: Prosodic evidence from right node raising in English. *Speech prosody, Aix-en-Provence*, 643–646.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P. et al. (1992). ToBI: A standard for labeling English prosody. In *Proceedings of the third international conference on spoken language processing* (pp. 867–870).

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 204–221. http://dx.doi.org/10.1037//0278-7393.26.1.204.

Steedman, M. (2000). Information structure and the syntax-phonology interface. *Linguistic Inquiry, 31*, 649–689. http://dx.doi.org/10.1162/002438900554505.

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1024–1037. http://dx.doi.org/10.1037/0278- 7393.25.4.1024.

Tullis, J. G., & Benjamin, A. S. (2010). On the effectiveness of self-paced learning. *Journal of Memory and Language, 64*, 109–118. http://dx.doi.org/10.1016/j.jml.2010.11.002.

Van Deemter, K. (1999). Contrastive stress, contrariety, and focus. In P. Bosch & R. van der Sandt (Eds.), *Focus: Linguistic, cognitive, and computational perspectives* (pp. 3–16). Cambridge, United Kingdom: Cambridge University Press.

Ward, P., & Sturt, P. (2007). Linguistic focus and memory: An eye movement study. *Memory & Cognition, 35*, 73–86. http://dx.doi.org/10.3758/BF03195944.

Watson, D., Gunlogson, C., & Tanenhaus, M. (2008). Interpreting pitch accents in on-line comprehension: H* vs L + H*. *Cognitive Science, 32*, 1232–1244. http://dx.doi.org/10.1080/03640210802138755.

Wolter, L., Gorman, K. S., & Tanenhaus, M. K. (2011). Scalar reference, contrast and discourse: Separating effects of linguistic discourse from availability of the referent. *Journal of Memory and Language, 65*, 299–317. http://dx.doi.org/10.1016/j.jml.2011.04.010.

Wright, D. B., Horry, R., & Skagerberg, E. M. (2008). Functions for traditional and multilevel approaches to signal detection theory. *Behavior Research Methods, 41*, 257–267. http://dx.doi.org/10.3758/BRM.41.2.257.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441–517. http://dx.doi.org/10.1006/jmla.2002.2864.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 162–185. http://dx.doi.org/10.1037/0033-2909.123.2.162.