

WHAT HAPPENED (AND WHAT DIDN'T):
PROMINENCE PROMOTES REPRESENTATION
OF SALIENT ALTERNATIVES IN DISCOURSE

BY

SCOTT FRAUNDORF

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Associate Professor Duane Watson, Chair
Associate Professor Aaron Benjamin
Professor Gary Dell
Associate Professor Kara Federmeier
Professor Elizabeth Stine-Morrow

ABSTRACT

In five experiments, I investigated how readers and listeners generate relevant contrasts in comprehending and remembering discourse. Past work has suggested that prominent words promote encoding of salient alternatives and that this benefits later memory, but it is unclear exactly which alternatives are considered or how consistent these benefits are across modalities and across individuals. Participants read or listened to discourses containing salient alternatives (e.g., *Malaysia* when the discourse also mentioned *Indonesia*). In Experiments 1 and 2, font emphasis in the initial presentation facilitated participants' ability to later reject the salient alternatives but not unmentioned items (e.g., *Portuguese scientists*), generalizing past effects of contrastive pitch accents. In Experiment 3, font emphasis facilitated rejections of salient alternatives but not less plausible alternatives that were nevertheless mentioned in the discourse. Online reading time measures in Experiment 2 indicated that emphasized words did not improve performance on all trials and only benefited memory to the extent that participants devoted extra time to them, although no such relation was observed in Experiment 3. The relationship of online reading time to later memory is consistent with views of language processing in which some aspects of linguistic representations may be left underspecified because they are time- or resource-consuming to generate. Further, the effortful processing of an alternative impaired memory for the rest of the discourse in populations with more restricted online processing abilities: older adults (Experiment 4) and younger adults who have lower scores on complex span scores (Experiment 5).

ACKNOWLEDGEMENTS

I would like to thank the members of my committee for their guidance, encouragement, and patience. Also thanks to the Communication and Language Lab and the Human Memory Cognition Labs at the University of Illinois for their comments and suggestions, Martha Fraundorf for assistance with stimulus construction, Shefali Khanna for recording audio stimuli, and Melissa Denis, Ran He, Kathleen Hourihan, Euil Lim, Laura Matzen, Gina Mulanlara, Sujin Park, Amie Roten, Kirstin Shafter, and Jonathan Tullis for data collection. I would also like to thank the financial support of National Science Foundation Graduate Research Fellowship 2007053221, a traineeship from National Institutes of Health Language Processing Training Program 5T32HD055272-12, and National Institutes of Health grants R01DC008774 to Duane Watson and R01AG026263 to Aaron Benjamin.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
CHAPTER 2: EXPERIMENT 1	16
CHAPTER 3: EXPERIMENT 2	30
CHAPTER 4: EXPERIMENT 3	46
CHAPTER 5: EXPERIMENT 4	61
CHAPTER 6: EXPERIMENT 5	78
CHAPTER 7: GENERAL DISCUSSION	88
REFERENCES	99
APPENDIX A: STIMULI FOR EXPERIMENT 2	118
APPENDIX B: TEST ITEMS FOR EXPERIMENT 2	123
APPENDIX C: STIMULI FOR EXPERIMENT 3	126
APPENDIX D: TEST ITEMS FOR EXPERIMENT 3	132

CHAPTER 1

INTRODUCTION

Linguistic material must not only be comprehended in the moment but, frequently, remembered later. And, because human memory frequently does not preserve a veridical representation of all possible material, linguistic form may be an important constraint on what or how much is actually remembered later.

For instance, one function of some elements of linguistic form, such as prosody, may be to lead readers or listeners to represent not only what *did* happen but also salient alternatives that did *not* happen. Fraundorf, Watson, and Benjamin (2010) found that certain prosodic cues in spoken discourse led listeners to encode information about a salient alternatives, such as *Scottish* as an alternative to *English* in (1b), and that this information helped them remember the events of the discourse. In particular, remembering that it was the Scottish knight who lost the tournament helped listeners later reject *Scottish* as the winner but did not affect rejections of an unmentioned item like *Welsh*, which was never 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 representing salient alternatives can contribute to memory for a discourse, it remains unclear how consistently such alternatives are encoded. For instance, it is unclear whether these effects require an explicit manipulation of prosody. In some accounts (Gussenhoven, 1983; Pierrehumbert & Hirschberg, 1990), the consideration of

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

The consistency with which alternatives are represented across instances and represented individuals is similarly unclear. The emphasis on *English* in (1b) may automatically lead listeners or readers to consider *Scottish* as a salient alternative. Alternately, readers may only sometimes consider this relationship and other times leave the alternative unspecified, especially if it is time-consuming to calculate. Moreover, if calculating an alternative is effortful, it is possible that the process of considering an alternative, such as recognizing that the *Scottish* knight did not win, could actually impair memory for other parts of a discourse, such as the fact that the knights held a jousting tournament in (1a).

A second unresolved issue in the processing of salient alternatives in discourse is how widely or narrowly the set of salient alternatives is defined. Fraundorf et al. (2010) compared the effects of contrastive prosody on a salient alternative, such as *Scottish* in (1a), to its effects on a completely unmentioned item. However, many discourses might contain items intermediate between these conditions: elements that were mentioned but are not necessarily salient alternatives. For example, the *Welsh knight* in (2a) is mentioned but is a less plausible alternative to *English* in (2b) than is *Scottish*. It is unclear whether prominence leads

comprehenders to consider all possible alternatives in a discourse or only those that are particularly likely alternatives.

(2a) The English and the Scottish knights held a jousting tournament, but the Welsh knight was sick and couldn't participate.

(2b) The English knight won the tournament.

In this dissertation, I examine three aspects of contrast in processing and remembering a discourse: (a) whether the representation of salient alternatives elicited by contrastive pitch accents could also be obtained in written discourse as a function of a different manipulation of prominence: font¹ emphasis such as italicization or capitalization, (b) whether the alternatives represented included *any* referent mentioned in the discourse or only those that were particularly likely alternatives to the true proposition, and (c) whether the mnemonic benefits of prominence are contingent on the online processing of the initial presentation of the discourse.

The Representation of Contrast in Discourse

Success in discourse comprehension may involve representing not only particular referents or propositions, but also calculating and representing a set of one or more salient alternatives (e.g., Rooth, 1992; Calhoun, 2009). For instance, Rooth (1992) has argued that placing a linguistic constituent in focus introduces a *focus semantic value*. The focus semantic value expands the semantic interpretation of a sentence by introducing a set of alternative

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). I follow the latter usage by referring to manipulations such as italicization as *font* changes.

propositions² that could have been formed had the focused element been replaced by something else. In some cases, the alternatives that comprise the focus semantic value may have been explicitly introduced into the discourse; in other cases, they may be pragmatically inferred. For example, (3) express the proposition that what maximizes cognition is elaborative rehearsal. Because the cleft structure focus *elaborative*, it also introduces a set of alternative propositions, such as *maintenance rehearsal maximizes cognition*, that could have been formed by replacing *elaborative* with other potential modifiers for *rehearsal*. That is, emphasizing *elaborative* highlights the fact that it is elaborative rehearsal, and not some other form of rehearsal, that maximizes cognition. Speakers and writers may introduce alternatives in this manner for any number of reasons, commonly including the desire to contrast two outcomes (Rooth, 1992).

(3) It's ELABORATIVE rehearsal that will maximize cognition.

More recent accounts (Calhoun, 2009) have proposed that this sort of contrastiveness is not a direct function of particular linguistic categories but occurs probabilistically: the more linguistically prominent³ a word is relative to expectations, the more likely it is to bring to mind a set of salient alternatives. Indeed, empirical results suggest that consideration of salient alternatives in online language processing may be brought about by any number of linguistic

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

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). Following Birch and Rayner (2010) and Calhoun (2009), I adopt the more general term *prominence* to refer to differences in importance or emphasis that may not necessarily be isomorphic to focus.

devices, including clefting (Sanford, Price, & Sanford, 2009), focus-sensitive particles such as *only* (Sedivy, 2002), many adjectives (e.g., *the tall glass* suggests a short glass is also present; Sedivy, Tanenhaus, Chambers, & Carlson, 1999) certain types of prosodic pitch accents (Ito & Speer, 2008; Watson, Tanenhaus, & Gunlogson, 2008), and font emphasis such as capital (or *Tall Man*) letters or italicization (McAteer, 1992).

The consequences of most of these devices for later representation of alternatives has been less frequently tested, but some evidence suggests consideration of alternatives can also 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 I term the *context passage*, established two pairs of items. For example, (4a) 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 facilitating memory for one pair would decrease memory for the other, but no such effects were observed.) A second part of the discourse, which I term the *continuation*, mentioned one member of each pair, as in (4b).

(4a) Both the British and the French biologists had been searching Malaysia and Indonesia for the endangered monkeys.

(4b) Finally, the British spotted one of the monkeys in Malaysia and planted a radio tag on it.

Each pair was used to establish a salient alternative to a proposition in the continuation. For example, the *Indonesia* and *Malaysia* in (4a) makes it plausible that the monkey could have been spotted in Indonesia, rather than Malaysia, in (4b). In the account proposed by Rooth (1992), the salient alternative *Indonesia* should form part of the focus semantic value if *Malaysia*

is made prominent in (4b).

In the original study by Fraundorf et al. (2010), the type of prominence that was manipulated was the type of prosodic pitch accent that listeners heard on each of the two critical words. The ToBI system for prosodic transcription (Beckman & Elam, 1997; Silverman et al., 1992) distinguishes between multiple types of prosodic pitch accents. 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).

The results of Fraundorf et al. (2010) suggested that the contrastive L+H* accent led participants to actively encode and remember a focus semantic value. A day after listening to the stories, participants took a recognition memory test that included three types of probes: correct statements, such as (5a), *alternative probes* that referred to items that were part of the original pairing in the discourse, such as (5b), and *unmentioned probes* that referred to items never present in the original discourse, such as (5c).

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

(5b) The French scientists found the endangered monkey.

(5c) The Portuguese scientists found the endangered monkey.

Both the contrast and unmentioned probes were false statements that were to be rejected, but they differed in whether or not the item to which they referred was likely to be part of the

focus semantic value: contrast probes referred to the salient alternative in the original discourse and should be part of a focus semantic value, while 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 facilitated rejection of the alternative probes, but, crucially, not the unmentioned probes. This suggests that the memory benefit from the L+H* accent was not simply due to improved encoding of the target item, which could facilitate rejections of *any* false probe. Rather, the effect likely stemmed from listeners representing something about the salient alternative (i.e., remembering that it was not the French scientists who found the monkey), which would benefit rejections only of that alternative and not of an unmentioned item.

To date, it is unclear whether these effects are unique to manipulations of spoken prosody. One account of the memory effects observed by Fraundorf et al. (2010) is that they reflect the meaning of the L+H* pitch accent. In some accounts, different intonational contours carry different meanings, and it is the particular meaning of L+H* to highlight or introduce salient alternatives (Pierrehumbert & Hirschberg, 1990; see Gussenhoven, 1983, for a similar argument in a somewhat different system of prosodic transcription). If consideration of alternatives is unique to the L+H* pitch accent, similar effects might not exist in situations without an explicit presentation of the L+H* accent, such as in reading written discourse. However, readers appear to generate implicit prosody when reading silently (Breen & Clifton, 2010; Fodor, 1998). Linguistic devices in written discourse, such as emphasizing a particular word with a font change, might thus also lead readers to implicitly generate a L+H* pitch contour and consider salient alternatives. Moreover, other accounts of prosody (e.g., Calhoun, 2009) have proposed that contrastive interpretations are not unique to particular pitch contours; rather, any information that is more linguistically prominent than expected highlights

alternatives. That account predicts that the representation of salient alternatives in memory should be observed across a wide range of manipulations, including font emphasis in written discourse.

Emphasis in Written Text

In the present study, I investigated whether font emphasis would, like contrastive pitch accents, lead to the representation of salient alternatives in memory. In explicit metalinguistic judgments, participants have described font emphasis as suggesting contrast (McAteer, 1992). To date, however, findings have been mixed as to whether font manipulations benefit comprehension. Emphasizing text using font changes has sometimes been observed to improve memory, including for confusable drug names (Filik, Purdy, Gale, & Gerrett, 2006) and for science texts (Golding & Fowler, 1992; Lorch, Lorch, & Klusewitz, 1995). In other cases, however, no benefit of font emphasis has been observed (Harp & Mayer, 1998). The hypothesis that font emphasis leads readers to encode a focus semantic value provides one possible explanation to these inconsistencies: remembering salient alternatives would benefit 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 type of emphasis used, and particular font changes may differ in their effectiveness or in their interpretation. For example, Filik et al. (2006) found that capital letters were more likely to benefit memory than colored text. McAteer (1992) elicited metalinguistic comparisons of capitalization and italicization and found that participants were more likely to assign a contrastive reading to italics than to capital letters, 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, I separately tested two different

manipulations: capital letters and italicization.

Reading Time and Depth of Processing

An additional benefit of assessing comprehension of written discourse is that participants' reading time provides a measure of their initial, online processing of the discourse. The representation of focus semantic values in memory may be contingent on the online processing of the emphasized material: Rooth (1992) has argued that fixing a focus semantic value is an optional process and not always performed. One plausible reason why readers might not always fix a focus semantic value is that this, and other aspects of language processing, may be time-consuming. The need to interpret a sentence in time 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). Whether particular items preferentially receive additional study time is thus an important determinant of later memory (Dunlosky & Connor, 1997; Tullis & Benjamin, 2010). It is possible that the degree to which readers invest time in calculating the focus semantic value in their original reading of 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 effort

(Ward & Sturt, 2007). Thus, it is unclear whether or not the representation of salient alternatives in memory would be predicted by readers' online processing of the original discourse.

Individual Differences in Contrast Processing

The possibility that calculating a focus semantic value may require additional, effortful processing predicts a dilemma: Although devoting extra processing to the focus semantic value can improve memory for the certain details, it may also diminish the resources available for encoding and remembering the rest of the discourse.

This tradeoff may be especially problematic for readers or listeners who already have diminished online processing abilities. For example, investigations of cognitive change across the lifespan have found that processing resources such as working memory and speed of processing decrease across the lifespan (Park, Lautenschlager, Hedden, Davidson, Smith, & Smith, 2002). This decline, along with gains in linguistic knowledge, is one of the major causes of age-related changes in discourse comprehension (Stine-Morrow, Miller, Gagne, & Hertzog, 2008). If older adults have fewer resources available to process a discourse, then effortful encoding of a focus semantic value is likely to impair their memory for the rest of the discourse.

Similarly, even within populations of the same age, there is variance in performing online tasks such as *complex span tasks*, in which participants must remember some material while also performing an online processing task. These demands are not unlike those involved in processing a discourse, in which new material must be processed while also being integrated with material held in memory (Graesser, Mills, & Zwaan, 1997). Consequently, complex span scores robustly predict individual differences in discourse comprehension (Daneman & Merikle, 1996). Although it is debated what underlying construct or constructs actually constrain performance on these tasks—individual differences in complex span performance have

alternately been attributed to processing speed (Salthouse, 1996), executive control (Engle, 2002), linguistic knowledge or skills (MacDonald & Christiansen, 2002; Farmer, Christiansen, & Kemtes, 2005), and inhibitory processing (Hasher, Zacks, & May, 1999)—there is little dispute that individual vary in their ability to meet the demands of complex span tasks. It is plausible, then, that young adults with low complex span scores may encounter a similar difficulty as older adults: encoding a focus semantic value may diminish their ability to encode and remember the rest of the discourse.

How Are Focus Semantic Values Defined?

A second issue in the calculation and representation of focus semantic values is which set of alternatives actually becomes the focus semantic value. Although it has been generally proposed that the alternative set is generally contextually defined (Rooth, 1992), context could define either a broad or narrow set of alternatives. For instance, one possibility is that comprehenders consider a broad set of alternatives, such as all the referents present in the discourse, or all the referents in a particular semantic class. This possibility is 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). Other experiments, however, suggest that the set of relevant alternatives in a discourse can be quite restricted. 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). Although these experiments have focused on how the *correct* referent of a referring expression is identified, they also suggest that the set of alternatives is relatively small.

The original experiments by Fraundorf et al. (2010) could not discriminate how strictly or narrower the set of alternatives was defined. In their recognition memory test, rejections of a highly salient alternative were compared to rejections of a completely unmentioned referent. As noted above, prosodic prominence increased rejections only of the salient alternative, suggesting its effect lay in encoding of the salient alternative in a focus semantic value. However, it is possible that this difference obtained either because the focus semantic value was fixed narrowly to the most salient alternative to the true proposition or because it was fixed more broadly to include *any* referent in the discourse, which would still exclude the completely unmentioned item.

One way of evaluating hypotheses of what is encoded in memory is by testing what false information can be rejected on the basis of those memories. This technique has been applied, for instance, in evaluating whether original memories are lost when later misleading information is encountered (McCloskey & Zaragoza, 1985). I tested the breadth of alternative sets by comparing rejection of likely alternatives to referents that were *mentioned* in the discourse but that was a less likely alternative. 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 a salient alternative (i.e., the photos taken of Saturn could have been lost instead). However, the discourse establishes that *Neptune* is a less likely alternative for (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 such as this provide an avenue for testing how broad a set of alternatives is considered by comprehenders. If only a limited set of alternatives are encoded in a focus semantic value, then font emphasis should benefit later rejections only of the most likely alternative (e.g., *Saturn* in the above example), and not of a mentioned but unlikely alternative (*Neptune*). However, if the focus semantic value is defined more broadly to encompass all referents given in the discourse, then rejections of both false statements should benefit from the focus semantic value.

Overview of Experiments

In five experiments, I investigated whether prominence in discourse led to the representation of salient alternatives consistently across manipulations of prominence and across comprehenders. I further tested how widely or narrowly this set of alternatives was defined. Experiments 1, 2, and 3 tested whether the representation of salient alternatives, previously observed in spoken discourse as a function of contrastive prosody (Fraundorf et al., 2010), could be obtained in written discourse as a function of font emphasis. Additionally, Experiments 2 and 3 tested whether online reading modulated the mnemonic benefits of contrastive emphasis. Experiment 3 then tested how wide or narrow a set of alternatives were considered by readers. Finally, Experiments 4 and 5 tested whether prominent information could impair memory for other, less prominent parts of a spoken discourse.

In Experiment 1, I first established that font emphasis could generate a contrastive

reading like that reported by Fraundorf et al. Both italics and capitals benefited memory in the same way that the contrastive (L+H*) pitch accents did in the Fraundorf et al. study: they facilitated rejection of contrast probes, but not of unmentioned probes, consistent with an account in which readers encode a focus semantic value in response to prominent information.

In Experiment 2, I then assessed online reading time as well as eventual memory to determine whether the emergence of a contrast effect in memory was mediated by initial reading time. There are multiple reasons why eventual memory might be unrelated to reading time. Fixing the focus semantic value might not be effortful and thus not modulate reading time. Alternately, fixing the focus semantic value might be effortful but routinely done upon encountering an emphasized word. This would lead to a consistent slowdown in reading time at the emphasized word, but, since this calculation happens uniformly across contexts, this slowdown would not be predictive of whether the emphasis benefits memory on a particular trial. However, if generating a focus semantic value is effortful, and readers do not always do so—perhaps because doing so would be time- or resource-consuming (Ferreira, 2003; Ferreira & Patson, 2007)—then emphasized words should only benefit memory to that extent that readers spend time to calculate the focus semantic value.

Next, in Experiment 3, I tested whether comprehenders consider a relatively wide or narrow set of alternatives to prominent information. I compared the effects of font emphasis on rejection of two types of false statements: a plausible alternative in the discourse, and a referent that was mentioned in the discourse but that was a less plausible alternative for the true proposition.

Finally, Experiments 4 and 5 tested how processing a focus semantic value could affect for memory for the discourse as a whole. Because the results of Experiment 2 suggested that

calculating a focus semantic value could sometimes be effortful and time-consuming, it is plausible that fixing a focus semantic value would take processing resources away from other discourse elements. Thus, the prominence of some information may decrease memory for the rest of the discourse, especially for populations with more restricted online processing resources. I tested this prediction with two such populations: older adults (Experiment 4) and young adults with lower scores on complex span tasks (Experiment 5).

CHAPTER 2

EXPERIMENT 1

I first tested whether font emphasis would benefit memory for discourse, and whether those benefits would lie in rejections of a salient alternative, as reported by Fraundorf et al. (2010) for contrastive (L+H*) pitch accents. Recall that in those, the key comparison was how a contrastive accent in discourses such as (4), reproduced below, affected later responses to three types of memory probes, reproduced below as (5).

(4a) Both the British and the French biologists had been searching Malaysia and Indonesia for the endangered monkeys.

(4b) Finally, the BRITISH spotted one of the monkeys in Malaysia and planted a radio tag on it.

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

(5b) The French scientists found the endangered monkey.

(5c) The Portuguese scientists found the endangered monkey.

The crucial comparison in this design is between probes that referred to a salient alternative in the original discourse, such as *French* in (5b), versus probes that referred to an unmentioned item, such as *Portuguese* in (5c). 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 evidence for the encoding of a focus semantic value, however, is whether the two types of probes differ in how they are

affected by prominence in the original discourse. One possibility is that the primary effect of prominence is to enhance encoding of the prominent item itself (e.g., that the British found the monkey). Enhancing memory for the correct proposition might also help readers reject the lures by process of elimination (the phenomenon of recollection rejection; Matzen, Taylor, & Benjamin, 2011; see also Brainerd, Reyna, and Estrada, 2006, for an example in the domain of discourse comprehension), but it should not apply exclusively to particular types of probes. That is, greater certainty that the British found the monkey should help reject both *French* and *Portuguese*. However, if the effect of prominence is to promote encoding salient alternatives in a focus semantic value (i.e., that the French did not find the monkey), then prominence should facilitate rejections only of the contrast lure and not the unmentioned lure. Knowing that the French did not find the monkey would not benefit rejections of *Portuguese*.

In Experiment 1, I 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 native English speakers at the University of Illinois participated in partial fulfillment of a course requirement.

Materials

Participants read 36 discourses, taken from Experiment 3 of Fraundorf et al. (2010). Each discourse began with a context passage, such as (4a) above, that established two contrasts, each between a pair of items. A subsequent continuation passage, such as (4b), mentioned one item from each of the contrast sets. 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. Emphasis was independently manipulated on each of the two critical words, so that, within a given passage, emphasis could be on the first, the second, both, or neither of the critical words, as in (7) below. The separate manipulation of two pairs per discourse was included for consistency with Fraundorf et al., 2010. As in their experiments, the young adult population as a whole did not show an effect of the properties of one pair on memory for the other pair. However, Experiments 4 and 5 will later show such effects in younger adults with lower scores on complex span tasks and in older adults.

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

The type of 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 (6a) above, referred to the correct item and should be affirmed. An *alternative probe*, such as (6b), referred to the other member of the pair in the original discourse and should be rejected. An *unmentioned probe*,

such as (6c), referred to an item from the same semantic category but that was never mentioned in the original discourse, and should also be rejected. Because there were two pairs per discourse, there were a total of 72 test items. Each participant saw only one probe for each item. 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 x 2 x 2 design: probe type (correct, contrast, or unmentioned) x presence of emphasis x emphasis type (capitals or italics). 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 a confound with syntactic position or 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 displayed to participants as a single paragraph; I separated them in the examples above purely for exposition. Participants took as

long as they wanted 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. However, accuracy might vary between true and false probes simply due to participants' overall tendency to respond *true* or *false*, rather than ability to discriminate one probe type from another. I instead analyzed participants' responses using the framework of the theory of signal detection (Green & Swets, 1996; Macmillan & Creelman, 2005; Wright, Horry, & Skagerberg, 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).

I 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 as a

function of predictor variables that can vary categorically (e.g., font emphasis) or continuously (e.g., reading time). I 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 determine 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 aggregation across all the 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), dichotomization of a continuous variable greatly reduces statistical power by discarding all the variation within each category (Cohen, 1983). Mixed effect logit models provide a natural way to analyze the relation of reading time to later memory because they model memory performance at the level of individual items and can easily incorporate continuously varying predictors such as reading time. Although reading time was not included as a predictor in Experiment 1, I apply the same methodology to Experiment 1 for consistency and easy comparison across experiments.

A secondary motivation for using mixed effects logit models is that in some conditions, particularly in Experiment 1, the proportion of *true* responses was low. Treating such proportions as the dependent variable in an ANOVA would be inappropriate in this case: 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 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 participants and items (propositions being tested). The fixed effects tested in Experiment 1 were the probe type, presence of emphasis in the original story, and emphasis type, as well as their interactions. All predictors were centered; doing so provides parameters corresponding to the main effects in an ANOVA analysis. The three levels of probe type were coded using two orthogonal contrasts: the first compared unmentioned probes to the other two probe types, and the second compared correct to alternative probes. All models were fit in the R Project for Statistical Computing using the *lmer* function of the *lme4* package (Bates, Maechler, & Bolker, 2011).

Proportion of *true* responses in each condition is displayed in Figure 1 as a function of probe type, presence of font emphasis in the original discourse, and emphasis type. Note that, in each discourse, there were two propositions tested, each of which could be independently analyzed. It was possible that memory for one contrast set (e.g., whether the *British* or *French* scientists found the monkey) could be influenced by whether or not the other contrast set (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., 1996) and on pitch accents (Fraundorf et al., 2010). (However, Experiments 4 and 5 will show differing results of this variable in other populations, including older adults). Consequently, I collapsed across this variable for all subsequent analyses in Experiment 1.

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 (presence of emphasis and probe type) did not improve the fit of the model, $\chi^2_{(20)} = 8.41, p = .99$. Random slopes by items for the effects of emphasis and probe type did improve the model, $\chi^2_{(20)} = 209.45, p < .001$, but no random slopes for emphasis type further improved the model (all $ps > .9$). Thus, I report results from the model with only random slopes by items for emphasis, probe type, and their interaction.

Parameter estimates for the final model are displayed in Figures 2 and 3. Overall, the odds of responding *true* rather than *false* were 0.61 (95% CI: [0.51, 0.73]), which reliably differed from chance, Wald $z = -5.57, p < .001$. This tendency to respond *false* is appropriate given that there were more false probes than true probes. The odds of responding *true* were 15.75 (95% CI: [10.22, 24.28]) times greater for correct and alternative probes than for unmentioned probes, $z = 12.49, p < .001$, indicating that participants were less likely to (incorrectly) affirm the unmentioned probes than an item mentioned in the discourse. The odds of responding *true* were 12.37 (95% CI: [5.86, 26.16]) times greater to correct probes than to alternative probes, $z = 6.58, p < .001$, indicating that participants were also able to successfully discriminate the true correct probes from the false alternative probes.

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*, $z = -.53, p = .60$. However, it did facilitate discrimination between the correct and alternative probes, with the odds ratio between correct and alternative probes being 5.12 times (95% CI: [2.08, 12.58]) greater when the critical word was originally emphasized, $z = 3.55, p < .001$. However, font emphasis did not reliably benefit discrimination between the unmentioned probes and other probes, $z = 1.47, p = .14$.

Finally, there was a 3-way interaction of probe type, presence of emphasis, and accent type, indicating that the benefit of emphasis in rejecting alternatives was stronger for capital letters than italicization, with the effect being 4.72 times (95% CI: [1.41, 15.84]) greater for participants who saw capitals rather than italics, $z = 2.51$, $p < .05$.

Discussion

Experiment 1 demonstrated an alternative-rejection effect similar to that observed by Fraundorf et al. (2010). Emphasizing one member of a pair of alternatives during the original presentation of a discourse improved the ability to reject the other alternative on a subsequent memory test, but did not benefit rejection of items that were never part of the alternative set. This suggests that the font emphasis led participants to encode a focus semantic value, or set of alternative propositions, which would help reject those alternatives but not items that were never in the set.

Experiment 1 extended 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), or with accounts in which any sufficiently prominent material is given a contrastive reading that brings to mind relevant alternatives (e.g., Calhoun, 2009).

Importantly, the 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 rejection of the unmentioned probes, suggesting the memory benefit lay in the encoding of a focus semantic value. Nevertheless, it is likely that the perceptual salience of emphasized text plays a role in its effects, and I return to this point in the General Discussion.

The findings of Experiment 1 are qualified by the somewhat high rate of rejections of the unmentioned probes even in the absence of emphasis, which may have masked any potential benefit of font emphasis in rejecting items that were never part of a discourse. In Experiment 3, I provide stronger evidence for representation of focus semantic values by demonstrating that font emphasis also fails to benefit rejections of a different type of probe that still does not mention the salient alternative but that elicits a substantially higher rate of false alarms.

Both capital letters and italicization promoted encoding of a focus semantic value in memory; in fact, the effect was stronger for capitalization. This conflicts with the finding that italicization is more apt to be judged to have a contrastive reading in offline metalinguistic judgments (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, 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 focus semantic value, or salient alternative, that could benefit 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, I assessed participants' online reading time and how it did or did not relate to participants' later memory.

Figures

Figure 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 italicization (bottom panel). Responding *true* is a hit to a correct probe and a false alarm to an alternative or unmentioned probe.

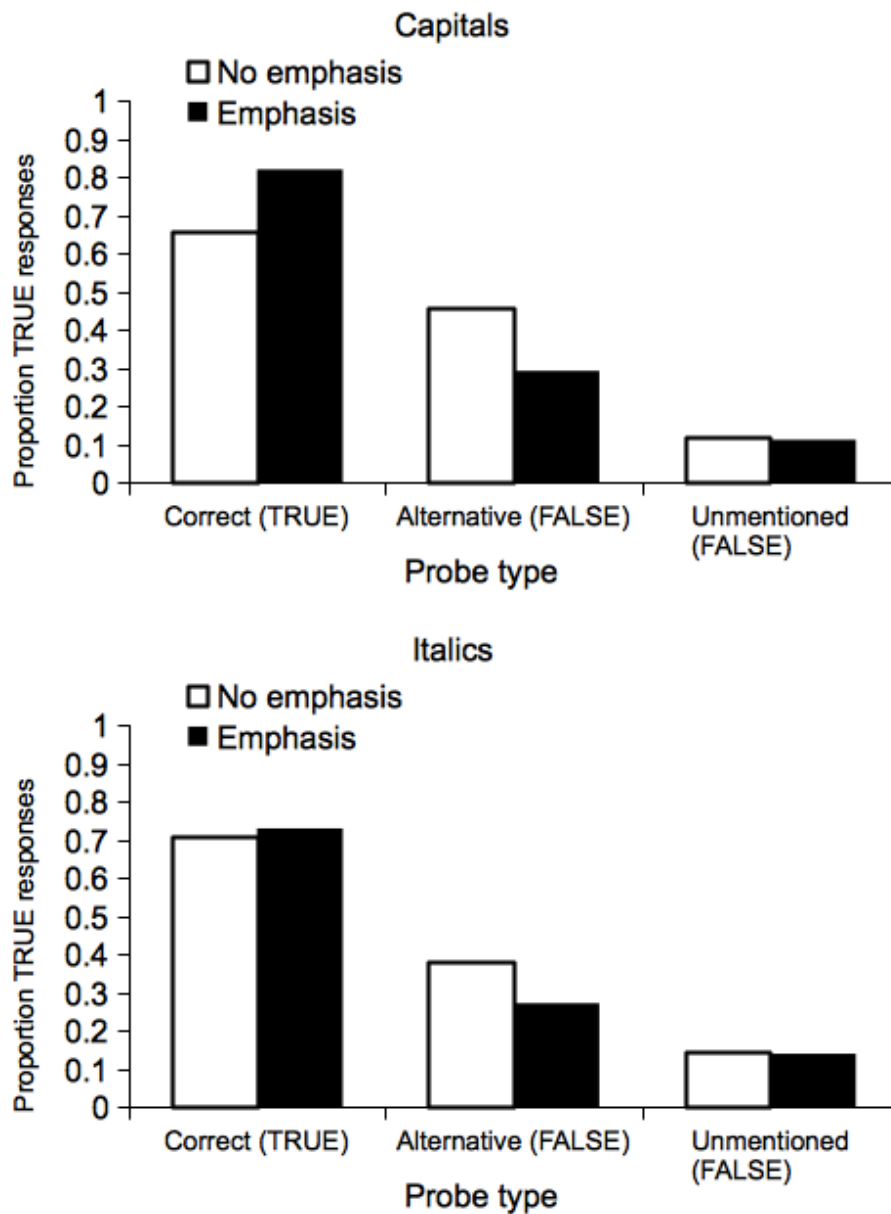


Figure 2: Fixed effect estimates for multi-level logit model of *true* responses in Experiment 1 (N = 1728, log-likelihood: -811).

Fixed effect	β	SE	Wald z	p
Intercept (response bias)	-0.50	0.09	-5.58	<.001
Correct/alternative probe vs. unmentioned (sensitivity)	2.76	0.22	12.49	<.001
Correct probe vs. alternative probe (sensitivity)	2.52	0.38	6.58	<.001
Emphasized word (effect on response bias)	-0.07	0.14	-0.53	.60
Capitalization vs. italics (effect on response bias)	0.27	0.27	0.99	.32
Emphasized word x capitalization (effect on response bias)	0.09	0.14	0.62	.53
Emphasized x correct/alternative vs. unmentioned (effect on sensitivity)	0.57	0.39	1.47	.14
Emphasized x correct vs. alternative (effect on sensitivity)	1.63	0.46	3.55	<.001
Capitalization x correct/alternative vs. unmentioned (effect on sensitivity)	0.56	0.31	1.80	.07
Capitalization x correct vs. alternative (effect on sensitivity)	-0.06	0.31	-0.19	.85
Emphasized word x capitalization x correct/contrast vs. unmentioned (effect on sensitivity)	0.48	0.62	0.77	.44
Emphasized word x capitalization x correct vs. alternative (effect on sensitivity)	1.55	0.62	2.51	.01

Note. SE = standard error.

Figure 3: 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
<i>Item</i>						
1. Intercept	0.18	—				
2. Emphasized word	0.04	-.44	—			
3. Correct/alternative vs. unmentioned probe	1.53	.14	-.76	—		
4. Correct vs. alternative probe	8.41	.05	-.07	.66	—	
5. Emphasis x correct/alternative vs. unmentioned	3.24	.47	-.62	.78	.63	—
6. Emphasis x correct vs. alternative	7.37	-.11	-.51	.36	-.15	.49
<i>Participant</i>						
1. Intercept	0.01					

CHAPTER 3

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 or not the memory benefit from the font emphasis was contingent on how participants originally read the emphasized words. One possibility is that calculating a focus semantic value in response to linguistic prominence may be time-consuming and only sometimes performed. 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. Alternately, however, fixing a focus semantic may require no extra time, or, even if it is time-consuming, readers may invariably invest the time to do so. Under either of these accounts, there would be no substantial variance in reading time related to calculating a focus semantic value that might predict later memory performance. As reviewed above, research to date has provided mixed results as to whether the depth of linguistic processing can be related to reading time, so it is unclear which of these possibilities would obtain.

Experiment 2 also used different discourses than Experiment 1, providing a further test of the generality of the memory benefits across materials.

Method

Participants

Forty-eight native English speakers at the University of Illinois 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 minutes allotted for the session and was

replaced with an additional participant.

Materials

The Experiment 1 materials were substantially modified 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 of alternatives (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 resolution 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 items in Experiment 1, the alternative proposition that would be expected in the focus semantic value 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 focus semantic value,

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 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. Emphasis type, the presence of emphasis, and probe 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, 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. I 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 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, in one discourse the referents mentioned in the context passage did not match the referents named in the continuation. I report results with data from this discourse omitted, but the inclusion or omission of this discourse 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 I examined reading times both on each critical word and on the *spillover* word that immediately followed. Mean reading time for each region and condition during the original presentation of the discourse is displayed in Figure 4. Because the reading times were positively skewed (skewness = 7.09) and thus non-normal, I used the natural log of the reading times (skewness = 0.92) as the dependent variable in my models. However, all of the effects reported were also reliable when the dependent variable was the untransformed reading time.

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 emphasis (emphasized or not), and emphasis type (capitalization or italics), resulting in a 2 x 2 x 2 design. Region and presence of emphasis were coded using dummy coding. This coding system first tests the simple main effects of 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 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, $\chi^2_{(9)} = 59.84, p < .001$, 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, $\chi^2_{(9)} = 31.23, p < .001$. 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, $\chi^2_{(26)} = 30.52, p = .25$.

Parameter estimates from the final model are displayed in Figures 5 and 6. First, consider the initial four parameters. These parameters test the simple main effects of emphasis within the critical word itself⁴. The model revealed that, overall, emphasized words ($M = 481$ ms) were read more slowly than words without emphasis ($M = 411$ ms), $t = 5.89, p < .001$. There was also an interaction with emphasis type; emphasis increased reading times more for participants who saw capital letters rather than italics, $t = 2.54, p < .05$.

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, emphasis, and emphasis type, with the difference between capitalization and

⁴ An alternate means of hypothesis testing in multi-level models with continuous dependent variables is Markov chain Monte Carlo simulations (Baayen, Davidson, & Bates, 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 because the t distribution has essentially converged to the normal distribution in models with thousands of observations.

italicization disappearing by the spillover region, $t = -2.44$, $p < .05$.

Memory

One goal of Experiment 2 was to assess whether the alternative-rejection effects in Experiment 1 varied as a function of participants' online reading time. That is, was a combination of both emphasis and increased reading time needed to obtain a benefit in rejecting the salient alternative? To test this hypothesis, I analyzed participants' memory performance using the same method as in Experiment 1, but 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 a focus semantic value. 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.

An alternative, then, is to examine only within-subject differences in reading time. I 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 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 I instead modeled log-transformed reading time.

I then analyzed the log odds of *true* responses as a function of probe type, presence of emphasis, and emphasis type, as well as residual reading time. Residual reading time was summed over the critical and spillover words; reading times more than three standard deviations from a participant's mean in that condition were replaced with the value three standard deviations from the mean, 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.

Mean rates of *true* responses in each condition are displayed in Figure 7, and parameter estimates from the model are displayed in Figures 8 and 9. A preliminary analysis indicated that, once initial reading time was accounted for, emphasis type (capitals or italics) made no further contribution to the model, $\chi^2_{(12)} = 11.21, p = .51$, so I dropped this variable to simplify the model. However, the model was improved by the inclusion of random slopes for probe type by participants, $\chi^2_{(5)} = 21.11, p < .001$, and by participants, $\chi^2_{(5)} = 425.70, p < .001$. No other random slopes approached significance (all $ps > .25$).

The memory effects observed in Experiment 1 were replicated. At a mean level of residual reading time, font emphasis facilitated discrimination between the correct and alternative probes: the odds of discriminating between correct and alternative probes were 1.55 times (95% CI: [1.02, 2.36]) greater for emphasized words, $z = 2.05, p < .05$. By comparison, font emphasis did not reliably benefit rejections of the unmentioned probe: emphasis only increased the odds of discriminating probes mentioned in the discourse from the unmentioned probes 1.02 times (95% CI: [0.69, 1.52]), $z = 0.11, p = .91$.

Additionally, Experiment 2 found that the benefit of font emphasis on rejecting

alternatives was amplified by increased reading time. The model revealed a three-way interaction between emphasis, probe type, and reading time on the rate of *true* responses. An increase of 1 log millisecond⁵ of reading time on emphasized material corresponded to a 2.33 times increase in the odds ratio between correct and alternative probes (95% CI: [1.07, 5.07]) greater, $z = 2.12$, $p < .05$. That is, font emphasis was more likely to facilitate rejection of the alternative probes the more readers slowed down on the emphasized words.

Again, this effect was limited to rejecting the salient alternatives. A 1-unit increase in residual reading time on emphasized material increased discrimination between unmentioned probes and other probes only 1.01 times (95% CI: [0.48, 2.14]), $z = 0.02$, $p = .98$. Moreover, the two-way interaction between reading time and probe type, collapsing across emphasis, was not significant, $z = -1.22$, $p = .22$. That is, increased reading time was not generally predictive of improved rejection of alternative probes. (In fact, the relation was numerically in the opposite direction.) Increased reading time only improved memory when it was spent on emphasized words.

There was one marginal effect on response bias: the odds of a *true* response were 1.17 times (95% CI: [0.98, 1.38]) greater for every 1-unit increase in residual reading time, regardless of whether a particular probe was true or false. This effect did not reach conventional levels of significance ($z = 1.77$, $p = .08$) and it is not clear what would account for it.

Discussion

Experiment 2 replicated the alternative-rejection effect in memory observed in Experiment 1. Emphasizing one member of a pair using font emphasis facilitated later rejections

⁵ 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.

of the salient alternative member of the pair. But, it did not benefit rejections of items that were unmentioned and unlikely to be part of the focus semantic value. 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 times increased predicted the extent to which font emphasis helped rule out 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. This pattern might suggest that calculating the focus semantic value is time-consuming: the degree to which readers spent extra time on the emphasized words predicted the extent to which the effect of the focus semantic value obtained in later memory.

Experiment 2 also clarifies the difference between the two types of emphasis tested in the present study: capitalization and italics. In Experiment 1, capital letters were observed to have a stronger benefit to memory 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 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 their effects on initial reading time.

The experiments thus far provide evidence that font emphasis can prompt readers to encode a focus semantic value, helping them to later reject certain salient alternatives. However, it is unclear exactly *which* alternatives are encoded in the focus semantic value. One possibility

is that the focus semantic value contains *any* alternative proposition possible in the discourse. Another possibility, suggested by findings that the set of alternatives considered in reference resolution can be tightly circumscribed (e.g., Brown-Schmidt & Tanenhaus, 2008), is that comprehenders consider only a restricted set of referents, encoding only those propositions that are particularly plausible or likely alternatives. 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 were neither salient alternatives nor mentioned in the discourse at all. Experiment 3 teased apart these possibilities by testing readers' rejections of items that had been mentioned in the discourse but which formed poor alternatives for the true proposition.

Figures

Figure 4: Mean reading time on target words (left panel) and spillover words (right panel) in Experiment 2 as a function of font emphasis and emphasis type.

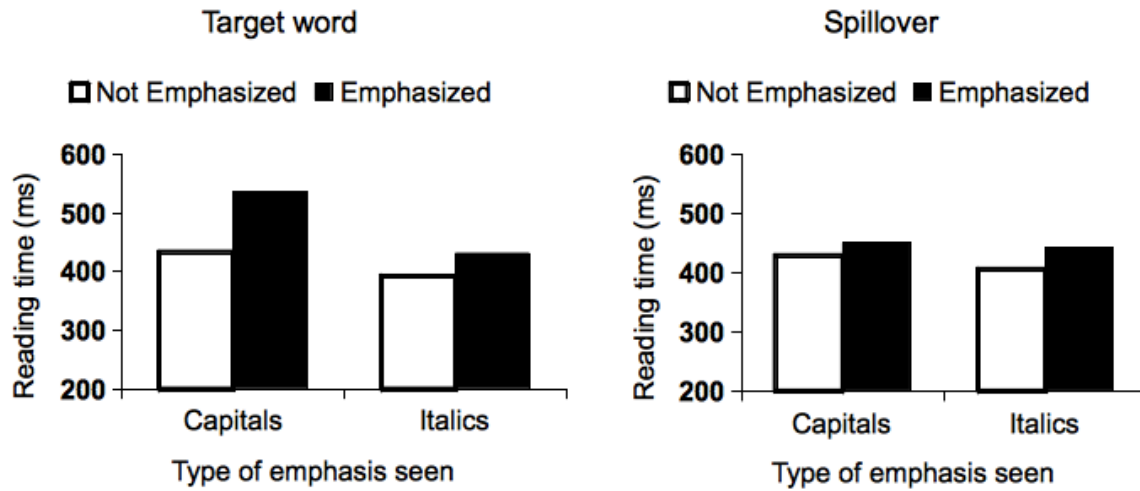


Figure 5: Fixed effect estimates for multi-level model of log reading time in Experiment 2

(N = 6720, log-likelihood: -2682).

Fixed effect	β	SE	<i>t</i>	<i>p</i>
Non-emphasized critical word (baseline)	5.90	0.04	138.10	<.001
Emphasized word	0.10	0.02	5.89	<.001
Emphasis type is capitalization (vs. italics)	0.06	0.08	0.66	.51
Emphasized word x emphasis type is capitalization	0.09	0.03	2.54	<.05
Spillover region	0.02	0.01	1.41	.16
Spillover region x emphasized word	-0.03	0.02	-1.48	.14
Spillover region x emphasis type is capitalization	< 0.01	0.02	-0.25	.80
Spillover region x emphasized word x capitalization	-0.10	0.04	-2.44	<.05

Note. SE = standard error.

Figure 6: 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
<i>Item</i>				
1. Non-emphasized critical word (baseline)	0.005	—		
2. Emphasized word	0.002	.31	—	
3. Spillover region	0.005	-.43	-.26	—
4. Spillover region x emphasized word	0.002	-.77	-.70	.02
<i>Participant</i>				
1. Non-emphasized critical word (baseline)	0.081	—		
2. Emphasized word	0.001	.72	—	
3. Spillover region	< 0.001	-.99	-.82	—
4. Spillover region x emphasized word	0.006	-.72	-.89	.79
<i>Residual variance</i>	0.122			

Figure 7: 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.

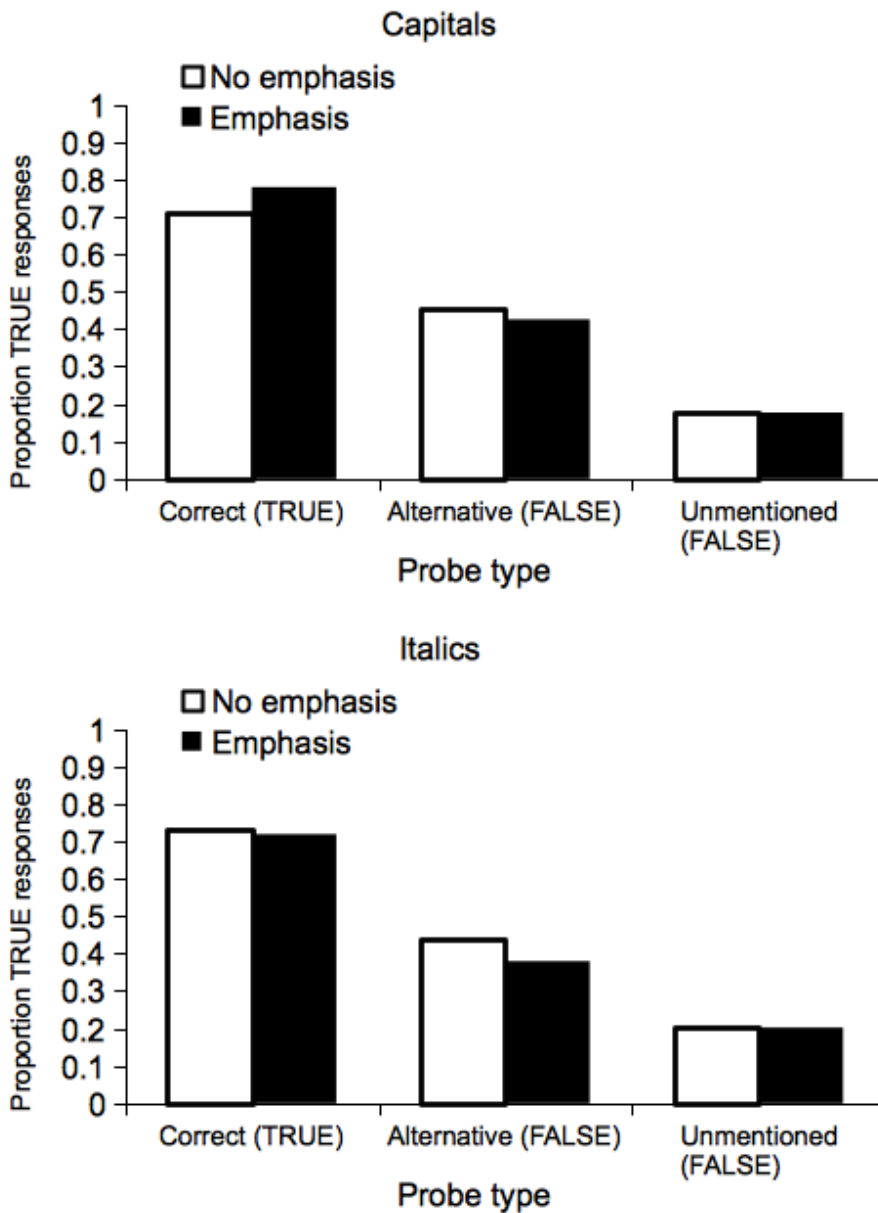


Figure 8: Fixed effect estimates for multi-level logit model of *true* responses in Experiment 2 (N = 3360, log-likelihood = -1674).

Fixed effect	β	SE	Wald <i>z</i>	<i>p</i>
Intercept (response bias)	-0.33	0.11	-2.87	<.01
Correct/alternative probe vs. unmentioned (sensitivity)	2.50	0.19	13.47	<.001
Correct probe vs. alternative (sensitivity)	1.58	0.33	4.75	<.001
Emphasized word (effect on response bias)	-0.01	0.09	-0.12	.90
Reading time (effect on response bias)	0.15	0.09	1.77	.08
Emphasized word x reading time (effect on response bias)	-0.03	0.17	-0.20	.84
Emphasized x correct/alternative vs. unmentioned (effect on sensitivity)	0.02	0.20	0.11	.91
Emphasized x correct vs. alternative (effect on sensitivity)	0.44	0.21	2.05	<.05
Reading time x correct/alternative vs. unmentioned (effect on sensitivity)	-0.12	0.19	-0.61	.54
Reading time x correct vs. alternative (effect on sensitivity)	-0.25	0.20	-1.22	.22
Emphasized x reading time x correct/alternative vs. unmentioned (effect on sensitivity)	0.01	0.38	0.02	.98
Emphasized x reading time x correct vs. alternative (effect on sensitivity)	0.84	0.40	2.12	<.05

Note. SE = standard error.

Figure 9: Summary of random item and participant effects and correlations in models of *true* responses in Experiment 2.

Random effect	s^2	Correlations	
		1	2
<i>Item</i>			
1. Intercept	0.29	—	
2. Correct/alternative vs. unmentioned probe	0.90	-.89	—
3. Correct vs. contrast probe	6.28	-.11	.31
<i>Participant</i>			
1. Intercept	0.30	—	
2. Correct/alternative vs. unmentioned probe	0.45	-.69	—
3. Correct vs. contrast probe	0.39	-.67	> .99

CHAPTER 4

EXPERIMENT 3

In Experiment 3, I examined whether comprehenders encode a narrow or broad set of alternatives in a focus semantic value. In particular, I tested whether font emphasis would facilitate rejections of poor alternatives to the true proposition that were nevertheless mentioned in the discourse. For example, consider context passage (6a), reproduced below. Both *Saturn* and *Neptune* are mentioned 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 Jupiter could have been lost instead.) However, *Neptune* is a poor alternative for *Jupiter* 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 and differentially related to reading time. Probe (8a) represents the alternative probe condition, the same as in prior

experiment, whereas probe (8b) involves an item that, while mentioned in the discourse, is a poor alternative for the true proposition. I term this new probe condition the *merely mentioned probe*. If prominence leads comprehenders to encode only the most salient alternatives in a focus semantic value, then font emphasis should benefit rejections only of the alternative probe and not of the merely mentioned probe. However, if comprehenders consider a broader set of alternatives—such as any referent that is instantiated in the discourse, or those in the semantic category—then emphasis may improve rejections of 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 confounding of probe conditions with lexical properties. In Experiments 1 and 2, the alternative and unmentioned probe conditions had 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 as 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, I 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 they invested additional time in reading 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 for the proposition in the continuation passage. For example, in example (6) above, the context passage establishes the pair *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 third 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. 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., *Nissan* and *Nick* 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. The true proposition was consistent across lists because it was of secondary importance; the primary interest was in the comparison between the two types of false probes.

The rotation of lexical items across lists introduced additional counterbalancing variables. To avoid a proliferation of experimental lists, I 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 controlling for initial reading time.

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 Figure 10 as a function of region and emphasis. I 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 I 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, $\chi^2_{(9)} = 83.67, p < .001$. Parameter estimates from the full model are displayed in Figures 11 and 12. 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 emphasis reliably vary across these regions.

Memory

Proportions of *true* responses are displayed in Figure 13 as a function of probe type and emphasis during the original discourse. As in Experiment 2, to assess whether later memory performance was predicting by online reading of the initial discourse, I calculated residual reading time 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). Reading times more than three standard deviations from the mean were replaced with the value three standard deviations from the mean, affecting less than 1% of the data.

I then modeled recognition memory decisions as a function of probe type, 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, $\chi^2_{(5)} = 51.72, p < .001$, and by items, $\chi^2_{(5)} = 92.94, p < .001$. No further slopes contributed reliably to the model, all p values $> .10$. Parameter estimates for the final model are displayed in Figures 14 and 15.

Overall, participants discriminated between the probe types. The odds of affirming a correct or alternative probe were 15.87 times greater (95% CI: [10.18, 24.74]), than those of affirming a merely-mentioned probe, $z = 12.21, p < .001$, and the odds of affirming a correct probe were 5.03 times greater than the odds of incorrectly affirming an alternative probe (95% CI: [3.67, 6.90]), $z = 10.04, p < .001$.

Font emphasis facilitated discrimination between the correct and alternative probes, as in previous experiments, with the odds ratio between correct and alternative probes being 2.17 times (95% CI: [1.40, 3.37]) greater for emphasized words than for non-emphasized words, $z = 3.47$, $p < .001$. However, emphasis did not reliably benefit discrimination of merely mentioned probes from the other probes. Although that discrimination was 1.22 times (95% CI: [0.85, 1.74]) better with font emphasis, the effect was not statistically reliable, $z = 1.07$, $p = .28$. Moreover, the model parameter estimates indicate that the effect of emphasis was nearly four times greater for discriminating correct and alternative probes (a 0.78 increase in the log odds) than for discriminating merely mentioned probes from the other probe types (a 0.20 increase in the log odds).

Unlike in Experiment 2, there were no effects of initial reading time on any aspects of participants' responding in the recognition memory task. In fact, increased reading time on emphasized words was numerically related to *worse* discrimination between correct and alternative probes, although this effect did not reach conventional levels of significance ($p = .12$).

Discussion

Experiment 3 replicated the benefit of font emphasis in rejecting salient alternatives to a true proposition on a recognition memory test. Experiment 3 also introduced a new type of memory probe referring to an item that was mentioned in the discourse but that the discourse had established as a poor alternative to the true proposition. Font emphasis conferred no benefit in correctly rejecting these probes. These results suggest that readers encoded only a narrow set of alternatives in response to the font emphasis. Had readers encoded every referent in the discourse as an alternative that did not happen, or encoded the true proposition in greater detail, font emphasis should have helped rule out any false alternative. Instead, the emphasis only

benefited discrimination between the true proposition and a particularly salient alternative to that true proposition.

The relation of reading time to memory performance diverged across experiments. In Experiment 2, longer reading time on emphasized words predicted discrimination between true statements and salient alternatives, suggesting the extra time was being used to encode a focus semantic value. Experiment 3 failed to replicate this benefit; indeed, reading time did not predict any aspect of performance on the recognition memory task. There are two possible reasons why reading time on emphasized words could be unrelated to discriminating correct and alternative probes: Calculating the focus semantic value might not have required any additional time. (Although readers did read the emphasized words more slowly, it is possible that this additional time was spent on processes such as decoding the capital letters rather than encoding a focus semantic value.) Or, readers might have consistently invested the time to encode the focus semantic value, such that variance in reading time did not index whether or not a focus semantic value was calculated on that particular trial.

Experiment 3 succeeded in eliminating the baseline differences between the types of false probes. As noted previously, the baseline rate of false alarms to the unmentioned probes in Experiments 1 and 2 was low. In those experiments, it was possible that font emphasis did not benefit rejections of the unmentioned probes not because it led readers to encode only a narrow set of alternatives, but simply because a floor effect obscured any benefit in rejecting other probes. In Experiment 3, however, 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 probes benefited from the font emphasis. This pattern provides strong evidence that font emphasis led to encoding of only a narrow set of alternatives and that

the difference between probes does not simply reflect a floor effect.

The comparison between the alternative and merely mentioned probes also rules out another alternate interpretation of the results of Experiments 1 and 2. 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 an item that had been mentioned and was more familiar in the context of the experiment, whereas the unmentioned probes referred to an item that had not been recently mentioned at all and, as a consequence, was also less familiar. Thus, it was possible that the differential effects of font emphasis on rejection of contrast and unmentioned probes in prior experiments reflected a difference between effects on familiar versus unfamiliar words rather than between items inside versus outside a focus semantic value *per se*. However, in Experiment 3, both the alternative and merely mentioned probes were mentioned and familiar within the context of the experiment. Nevertheless, font emphasis selectively facilitated rejections only of the alternative probes, suggesting its benefit lay in encoding of that alternative in a focus semantic value.

The experiments thus far provide consistent evidence that prominent information in a discourse can lead to the encoding of a narrow set of salient alternatives to the prominent material. However, the evidence has been inconsistent as to whether this process requires additional online processing: Experiment 2 found that the mnemonic effects of prominent words increased with additional, effortful processing, while Experiment 3 did not. An additional source of evidence that can inform whether online processing mediates representation of salient alternatives is potential differences between individuals in how they make use of prominence. Some individuals may be more or less apt to selectively devote extra resources to important

information. Or, for individuals with restricted online processing ability, calculating a focus semantic value may decrease their ability to encode the rest of the discourse. I tested whether either of these patterns would be observed in older adult listeners in Experiment 3, and in younger adults with lower complex span scores in Experiment 4.

Figures

Figure 10: Mean reading time in Experiment 3 on target words and spillover words as a function of font emphasis.

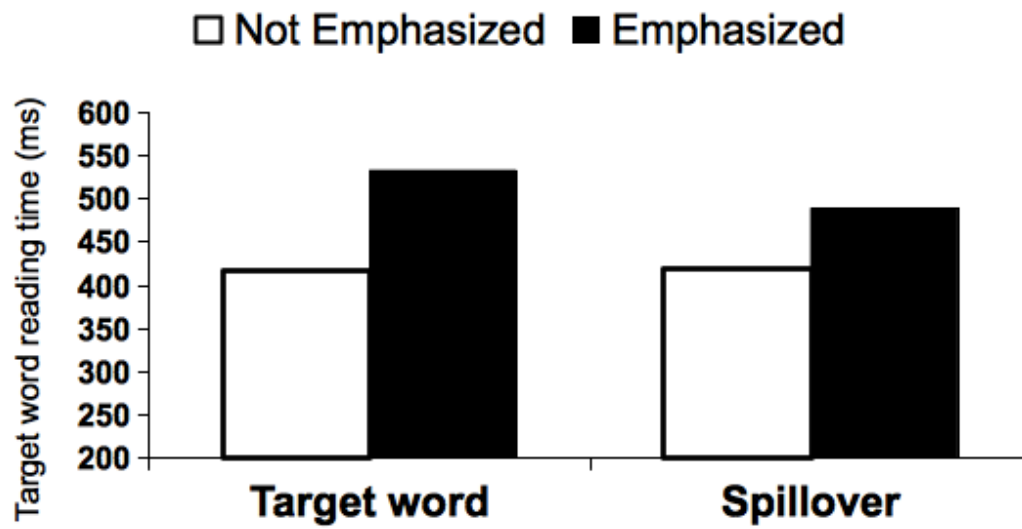


Figure 11: Fixed effect estimates for multi-level model of log reading time in Experiment 3

(N = 6912, log-likelihood: -3281).

Fixed effect	β	SE	<i>t</i>	<i>p</i>
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 x emphasized word	-0.02	0.03	-0.79	.43

Note. SE = standard error.

Figure 12: 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
<i>Item</i>				
1. Non-emphasized critical word (baseline)	0.002	—		
2. Emphasized word	0.011	-.24	—	
3. Spillover region	0.005	.50	-.14	—
4. Spillover region x emphasized word	0.008	.21	.96	.36
<i>Participant</i>				
1. Non-emphasized critical word (baseline)	0.094	—		
2. Emphasized word	0.021	.72	—	
3. Spillover region	0.001	-.78	-.64	—
4. Spillover region x emphasized word	0.026	-.81	-.83	.96
<i>Residual variance</i>	0.140			

Figure 13: 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.

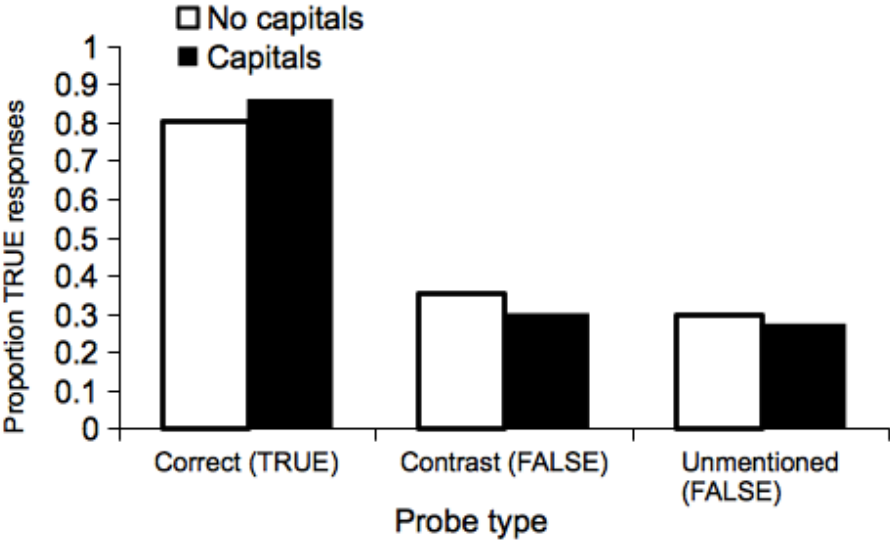


Figure 14: Fixed effect estimates for multi-level logit model of *true* responses in Experiment 3 (N = 3456, log-likelihood = -1805).

Fixed effect	β	SE	Wald z	p
Intercept (response bias)	-0.06	0.09	-0.62	.54
Correct/alternative probe vs. merely mentioned (sensitivity)	2.77	0.23	12.21	<.001
Correct probe vs. alternative (sensitivity)	1.62	0.16	10.04	<.001
Emphasized word (effect on response bias)	-0.02	0.09	-0.18	.86
Reading time (effect on response bias)	0.02	0.07	0.28	.78
Emphasized word x reading time (effect on response bias)	-0.18	0.15	-1.24	.21
Emphasized x correct/alternative vs. merely mentioned (effect on sensitivity)	0.20	0.18	1.07	.28
Emphasized x correct vs. alternative (effect on sensitivity)	0.78	0.22	3.47	<.001
Reading time x correct/alternative vs. merely mentioned (effect on sensitivity)	0.13	0.15	0.87	.38
Reading time x correct vs. alternative (effect on sensitivity)	-0.18	0.18	-0.98	.33
Emphasized x reading time x correct/alternative vs. merely mentioned (effect on sensitivity)	0.26	0.30	0.88	.38
Emphasized x reading time x correct vs. alternative (effect on sensitivity)	-0.56	0.36	-1.54	.12

Note. SE = standard error.

Figure 15: Summary of random item and participant effects and correlations in models of *true* responses in Experiment 3.

Random effect	s^2	Correlations	
		1	2
<i>Item</i>			
1. Intercept	0.34	—	
2. Correct/contrast vs. merely mentioned probe	0.74	-.41	—
3. Correct vs. alternative probe	1.48	-.48	.39
<i>Participant</i>			
1. Intercept	0.07	—	
2. Correct/contrast vs. merely mentioned probe	0.32	-.94	—
3. Correct vs. alternative probe	0.84	-.76	.94

CHAPTER 5

EXPERIMENT 4

Experiments 2 and 3 provided mixed evidence as to whether encoding a focus semantic value in response to prominent information required additional processing time. This issue has implications for how prominence might be understood by older adults—and, in turn, how prominence is used by older adults can inform whether or not encoding a focus semantic value is time-consuming. Accounts of cognitive aging have proposed that older adults differ from younger adults in multiple ways that bear on the encoding of focus semantic values: they may be either more selective or less selective at preferentially encoding important or difficult information, they typically have greater linguistic knowledge, and they typically have more limited online processing ability.

In Experiment 4⁶, I compared how young and older adults encode prominent information in a discourse. One issue in experimental investigations of cognitive aging is that age differences may be exaggerated by tasks that are less naturalistic (Benjamin, 2010; Castel, 2008). For instance, Tullis and Benjamin (in press) have argued that strategy use in older adults is equivalent or superior to that of younger adults when the strategy in question reflects naturalistic demands on memory. Because the moving window reading paradigm used in prior experiments is not a particularly naturalistic task, in Experiment 4, participants simply listened to spoken

⁶ This chapter was adapted from previously published material. Copyright © 2012 by the American Psychological Association. Adapted with permission. The official citation that should be used in referencing this material is 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. doi: 10.1037/a0024138. The use of this information does not imply endorsement by the publisher. I thank my co-authors, Duane Watson and Aaron Benjamin, for their assistance with the design of the research and the preparation of the manuscript.

discourse. As in Fraundorf et al. (2010), I manipulated whether the critical words received a presentational (H*) or a contrastive accent (L+H*).

The effects of contrastive accents might differ between older adults and young adults in at least two ways. First, it has been proposed (Rooth, 1992) that fixing a focus semantic value is not obligatory. Older adults might be more or less apt than young adults to devote the additional resources to contrastively accented items needed to encode a focus semantic value. Second, if computing a focus semantic value is time-consuming, encoding a focus semantic value for one item could diminish opportunities to encode the *rest* of the discourse, and do so especially for older adults. I tested whether older adults would differ from young adults in either of these aspects, and further considered whether any age-related changes reflected differences in linguistic knowledge or in online processing ability.

Age Differences in Encoding Selectivity

Age differences in episodic memory have sometimes been attributed to differences in the ability to select information for further processing (e.g., Dunlosky & Connor, 1997; Healey, Campbell, & Hasher, 2008). One account is that older adults are not as effective as young adults at selectively attending to the most important or difficult items, and this may cause the age-related decline in episodic memory typically observed in laboratory tasks.

For instance, Dunlosky and Connor (1997) found that older adults were less apt than young adults to selectively devote additional study time to those cue-target pairs that they had answered incorrectly on a previous cued recall test, even when information about their past performance was presented along with the item. In that case, differences in selectivity accounted for the majority of age-related variance in recall performance.

Differences in selectivity have also been observed in the domain of discourse

comprehension. For example, older adults are sometimes less apt than younger adults to discriminate more important from less important propositions in their recall of a text, especially as the task becomes more demanding (e.g., Dixon, Hultsch, Simon, & von Eye, 1984; Hartley, 1993; Stine & Wingfield, 1988). Although these effects have been most frequently tested in reading of written text, they have also been observed in comprehension of spoken discourse. For instance, Titone, Prentice, and Wingfield (2000) tested resource allocation using the auditory moving window paradigm, in which participants self-paced delivery of segments of spoken speech. They found that, compared to young adults, older adults' allocation of study time was less sensitive to syntactic boundaries and to discourse importance. Self-pacing of encoding also benefited older adults less than young adults on a later recall test for the material. These results suggest that older adults were less apt to preferentially allocate encoding resources to the most important or difficult materials, and that this difference led to a deficit in memory for the discourse. This view—that older adults are generally less strategic in their discourse comprehension—suggests that older adults should be less apt to select specific items on the basis of contrastive accenting and encode focus semantic values for them.

An alternate view of memory aging, however, is that selectivity *increases* over the lifespan. For instance, Castel (2008) has argued that, because older adults both have greater knowledge about what is important and may perceive their resources as more limited, they are more apt to focus on high value information. Consequently, it is in *less* prominent information where age differences should be greatest: older adults ignore less prominent material to focus on the prominent material, whereas young adults make an effort to remember everything.

This theory is supported by cases in which older adults appear to be as or even more selective than younger adults in memory tasks. Castel, Benjamin, Craik, and Watkins (2002)

tested recall of word lists in which each word was presented with a number during encoding. This number determined how many points participants received for recalling the word during a subsequent test phase. In this paradigm, older adults were just as likely as younger adults to recall the most valuable (high point value) words, although young adults were more apt to remember some less valuable items in addition. Older adults actually outperformed young adults on a measure of how well participants optimized their selection of high-value words given the total number of items they recalled. In the domain of discourse comprehension, Dixon et al. (1984) found that adults with higher verbal ability, as assessed by a vocabulary test, were just as successful as younger adults at remembering the most important propositions in a written discourse. Age deficits emerged only for less important, subordinate details.

These results indicate that, in some situations, older adults can be as effective as young adults in remembering valuable information and it is in less prominent information that age differences in memory emerge. This value-directed processing account suggests that older adults may be just as likely as young adults, or even more likely, to selectively encode additional information about contrastively accented items.

Age Differences in Processing Limitations

A second, related way that older adult listeners may differ from younger adult listeners is that young adults typically outperform older adults on tasks requiring fluid online processing (Park et al., 2002; Salthouse, 2004), such as perceptual comparisons (Park et al., 2002; Salthouse, 1996, 2004; Stine-Morrow et al., 2008) or working memory (Park et al., 2002; Salthouse, 2004; Stine-Morrow et al., 2008). Indeed, a decrease in online processing speed has been suggested to be one reason that older adults may be more selective in what they encode (e.g., Castel 2008) and has been a major component of models of how language processing

changes with age (Stine-Morrow et al., 2008).

These processing limitations may have consequences for how focus semantic values are encoded in response to prominent information. If older adults are more restricted in their online processing, they may either have insufficient time to encode a focus semantic value, or doing so may leave them without sufficient opportunity to encode the *rest* of the discourse. Although older adults can often compensate for processing declines by allocating additional resources to difficult material (Stine-Morrow et al., 2008), they may be less apt to do so when listening to spoken discourse, for which there is less control over the presentation rate. Thus, whether or not older adults show any penalty as a result of processing prominence provides an additional source of information as to whether or not it is time-consuming to encode a focus semantic value.

Use of Prosody by Older Adults

Finally, it should be noted that some work has specifically tested how older adults comprehend prosody. Older adults have been less successful than younger adults at identifying the emotional and attitudinal information conveyed by speakers' prosody (Orbelo, Testa, & Ross, 2003; Orbelo, Grim, Talbott, & Ross, 2005) to a greater extent than would be expected given age-related changes in audition (Orbelo et al., 2005). Use of the linguistic information conveyed by prosody, however, appears to be well-preserved across the life-span. Older adults and younger adults make similar use of lexical stress in identifying words (Wingfield, Lindfield, & Goodglass, 2000) and of prosodic boundaries in resolving syntactic ambiguities (e.g., Kjelgaard, Titone, Wingfield, 1999; Titone, Koh, Kjelgaard, Bruce, Speer, & Wingfield, 2006).

Changes with age in the comprehension of pitch accents, however, have been less studied, and the results have been mixed. Kemper and Harden (1999) found that exaggerated use of contrastive accents actually *impaired* older adults' online comprehension of map directions.

By contrast, Stine and Wingfield (1987) found that the presence of prosodic boundaries and pitch accents benefited older adults' recall of sentences more than younger adults', but this manipulation did not test pitch accents separately from prosodic boundaries. Similarly, Cohen and Faulkner (1986) found that the presence of pitch accents in spoken news stories improved cued recall performance more for older adults than it did for young adults. However, in both of those experiments, the presence or absence of pitch accents was manipulated across the passage as a whole, so it is unclear how pitch accents affected the allocation of attention to specific information within the passage.

Present Study

I compared young and older adults in how contrastive pitch accents affected their memory for discourse. Participants listened to discourses such as (3), reproduced below as (9). These discourses had the same structure as those in Experiments 1 and 2; that is, they contained two pairs of two items each, and the item to be used in the third probe condition was wholly unmentioned. However, the discourses were presented in the auditory modality, and, rather than manipulating font emphasis on the critical words, I manipulated whether the critical words were heard with a presentational (H*) pitch accent or a contrastive (L+H*) accent.

(9a) Both the British and the French biologists had been searching Malaysia and Indonesia for the endangered monkeys.

(9b) Finally, the British spotted one of the monkeys in Malaysia and planted a radio tag on it.

Experiments 1 through 3 used multiple types of memory probes in order to test exactly how prominence altered participants' memory for the discourse. Because these experiments

already established that the benefit lay in rejections of the salient alternative and because the purpose of Experiment 4 was instead to investigate age-related changes in the allocation of resources to encoding, I used a simpler memory task in Experiment 4 to reduce the number of cells in the design. Participants were presented with a two alternative forced-choice task, in which they simply had to indicate which of the two items in each pair (e.g., *British* or *French*) had appeared in the continuation.

Contrastive accents could affect performance on this memory task in two ways. First, a contrastive accent on one word might facilitate memory for that pair. For instance, a contrastive accent on *British* in (9b) might improve memory for the *British* versus *French* distinction. I term this possibility the *accent boost*. Second, a contrastive accent on one detail might direct attention or resources away from *other* information in the story. For instance, a contrastive accent on *British* in (9b) might impair memory for the proposition that the monkey was found in *Malaysia*. I term this possibility the *other-accent penalty*. These effects are not mutually exclusive. The accent boost likely reveals the effects of pitch accenting on promoting the encoding of a focus semantic value, whereas the other-accent penalty will only be revealed if insufficient resources are available to encode other information after encoding the focus semantic value.

Among young adults, Fraundorf et al. (2010) found the accent boost but no other-accent penalty. That is, contrastive accents improved memory for an accented item but did not impair memory for other propositions in the discourse. The same pattern was also obtained for young adults reading font emphasis in Experiments 1, 2, and 3.

In Experiment 4, I compared the effect of contrastive accents in this task for older adults and young adults. If older adults are less apt to select particular information to encode with more depth, then contrastive accents—which may lead listeners to additionally encode a focus

semantic value—should have smaller effects. However, if age differences are actually greater for *less* emphasized details, then older adults should show similar or greater memory for the important, contrastively accented details. In addition, older adults may be more apt to show the other-accent penalty; that is, when a contrastive accent makes one word particularly prominent, older adults would be less apt to encode the others.

Method

Participants

Forty-eight undergraduate students at the University of Illinois participated for partial fulfillment of a course requirement. Forty-eight community-dwelling adults (age range: 60 to 80 years; $M = 68$, $SD = 6.5$) were recruited through advertisements in campus publications and participated for a cash honorarium. All participants in both groups were native speakers of American English and all of the older adult participants scored at least 27 of 30 on the Mini Mental State Exam (Folstein, Robins, & Helzer, 1983).

The older adults had a mean score of 35.2 ($SD = 3.0$) on the Shipley Institute for Living Vocabulary Scale (Shipley, 1940). Vocabulary scores were not collected from the present sample of young adults, but a prior sample of 25 young adults from the same population had a mean score of 29.4 ($SD = 4.1$). Thus, there was some evidence that older adults had greater vocabulary knowledge, as is typically observed (Park et al., 2002; Stine-Morrow et al., 2008).

Orbelo et al. (2005) have shown that older adults' comprehension of prosody is not predicted by hearing sensitivity beyond the ability to hear the speech stream. Consequently, I did not test participants' hearing beyond their ability to hear the recorded stories. Participants in both age groups were able to adjust the volume of the computer task to ensure that the stories were audible.

Materials

48 short recorded discourses from Fraundorf et al. (2010) were used, taking the same form as those in prior experiments. The discourses were recorded by a female research assistant with an Inland Northern American English accent (Labov, Ash, & Boberg, 2006), appropriate for the region.

In each story, the type of pitch accent on each critical word in the continuation was orthogonally manipulated across participants between a presentational accent (H*) or a contrastive (L+H*) accent. Thus each story could be heard with a contrastive accent on the first contrast set, on the second contrast set, on both, or on neither.

To ensure that the stimuli differed only in the pitch accents on the target words, the different tokens of the critical word were spliced into a carrier sentence that did not vary across conditions. I administered a post-experiment survey to verify that the splicing did not result in stimuli that sounded unnatural. None of the participants in the two experiments that used the recorded materials (Experiments 4 and 5) noticed the splicing.

Acoustic analyses confirmed that the words with L+H* had greater mean F_0 , duration, and intensity, consistent with past descriptions of contrastive accents (Selkirk, 2002). These effects obtained both when the measurements were made on just the syllable carrying primary word stress, where pitch accents are argued to be realized (e.g., Ladd, 2008), and on the entire word.

The complete list of materials and further details on the acoustic measurements are available in Fraundorf et al. (2010).

Procedure

Participants first listened to a sample recording of the speaker and adjusted the computer

volume until they could easily hear the recording. The experiment began with a study phase in which participants listened to all 48 stories, presented in random order. During this time, the computer screen was blank. There was a 5 s delay between stories. After 24 of the stories had been presented, the computer informed participants they were halfway through the study phase and allowed them to take a break before continuing.

After participants had listened to all 48 stories, they proceeded to a test phase. During the test phase, each discourse was presented in text form, with the two critical words in the continuation replaced by blanks, as in (10). Participants did not hear the stories during the test phase and hence received no prosodic information during test.

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

Memory was tested one critical word at a time. The two items from the pair in the original discourse (e.g., *British* and *French*) were displayed on screen and participants chose one of them with a key press. The discourses were presented in the same order as during the study phase, with a 500 ms delay between the tests of each critical word and a 1000 ms delay between discourses.

After the test phase, participants completed a structured debriefing questionnaire in which they were asked whether they had heard anything odd or unusual in the recordings. No participant reported anything that suggested they had detected the splicing.

Results

Accuracy of recognition memory was analyzed as a function of three factors: the accent on the critical word being tested, the accent on the other critical word in the continuation, and age. Mean accuracy in each condition is displayed in Figure 16.

The model included fixed effects of age group, the accent on each target (H* or L+H*), and the interactions of these factors. A random slope of age by item improved the fit of the model in a likelihood ratio test, $\chi^2_{(2)} = 31.68, p < .0001$, indicating the age difference was larger for some items than others. The model was also marginally improved by a random slope of target accent by participants, $\chi^2_{(2)} = 5.19, p = .07$. Because this random slope did not reach conventional levels of significance, and because the amount of variance in this slope was extremely small ($< .03$), I report results from the model without it, but all reported effects were reliable both with and without this slope. No other random slopes contributed reliably to the model.

Parameter estimates for the final model are displayed in Figures 17 and 18. The accent placed on a referent reliably affected memory; the odds of correct recognition for words receiving a contrastive accent ($M = 85\%$) were 1.82 times greater (95% CI = [1.54, 2.14]) than for words receiving a presentational accent ($M = 77\%$), consistent with the results of Fraundorf et al. (2010).

Age did not reliably interact with the accent placed on a particular item, Wald $z = -0.53, p = .60$, n.s. That is, both younger and older adults showed an equivalent benefit to memory from a contrastive accent. However, age interacted with the accent placed on the *other* critical word in the story. For older adults, targets were less likely to be remembered if the other critical word had a contrastive accent ($M = 81\%$) than if it did not ($M = 83\%$); odds of recognition in this condition were 0.77 times lower (95% CI = [0.61, 0.98]) for older than for young adults. There

was no evidence of such an effect for young adults; in fact, young adults' memory was numerically *better* if the other contrast set received a contrastive accent ($M = 81\%$) than if did not ($M = 79\%$).

The overall effect of age on recognition was not reliable, $z = 0.91$, $p = .36$.

Discussion

In Experiment 4, contrastive accents were observed to facilitate memory, and to do so to an equivalent degree for both young and older adults. This indicates that older adults do not show a reduced ability to use contrastive accents to facilitate memory. This finding is consistent with other work showing that other linguistic uses of prosody, such as syntactic processing (Kjelgaard et al., 1999; Titone et al., 2006) and word recognition (Wingfield et al., 2000), also remain well preserved across the life span.

If anything, older adults showed evidence of *greater* sensitivity to pitch accents. When one critical word was heard with a contrastive accent, older adults appeared to prioritize that item to the detriment of others. For instance, a contrastive accent on *British* impaired older adults' memory for the *Malaysia/Indonesia* distinction. Young adults, consistent with Fraundorf et al. (2010), did not show this effect. To the extent that contrastive accents indicate that certain information should be prioritized for the encoding of a focus semantic value, these results are broadly consistent with the results of Castel et al. (2002), who found that older and young adults were equally successful at remembering at high value information, but older adults were less successful than young adults at remembering less valuable information.

Why do older, but not young, adults show this other-accent penalty? Castel (2008) has argued that selectivity may increase across the lifespan for multiple reasons, including both increased knowledge of what information in the world is important and limitations in processing

resources. This proposal is consistent with general studies of cognitive change across the lifespan, which have found that online processing abilities such as working memory and speed of processing decrease across the lifespan, whereas verbal knowledge increases (Park et al., 2002). It is also in accord with models of discourse comprehension (Stine-Morrow et al., 2008) that model age-related changes in reading time as a function of decreases in processing resources and increases in verbal ability.

Either of these changes could potentially drive the other-accent penalty observed in Experiment 4. For example, limited processes resources could force older adults to encode only the most important details. This is consistent with the results of Experiment 2, which suggested that encoding a focus semantic value was time-consuming. Encoding a focus semantic value for an item with a contrastive accent would leave older adults with less time to encode the rest of the discourse, creating an other-accent penalty. This *limited resources hypothesis* predicts that an other-accent penalty might also be evident in some young adults who are also restricted in online processing resources such as working memory ability.

However, Experiment 3 found no relation of initial processing time to later memory. And, indeed, an alternate explanation for the results of Experiment 4 that does not require changes in processing time is age differences is that older adults are likely to have more experience with the distribution of contrastive accents. As noted above, linguistic knowledge such as vocabulary increases over the lifespan (Park et al., 2002; Stine-Morrow et al., 2008), and, in Experiment 4, the older adults had a higher mean vocabulary score than the young adult population. This increased linguistic knowledge may make older adults better attuned to contrastive accenting as a cue indicating that a focus semantic value should be encoded. They might prioritize contrastively accented information for encoding even if they would have

resources to encode additional information. This *linguistic knowledge hypothesis* does not predict that limited resources underlie the other-accent penalty. If anything, a relationship with working memory might obtain in the opposite direction: young adults who score higher on working memory tasks typically show greater sensitivity to other linguistic cues such as verb distributional statistics (Pearlmutter & MacDonald, 1994), although the reasons for this relationship remain debated (MacDonald & Christiansen, 2002).

In Experiment 5, I pit these hypotheses against each other by testing young adults who varied in their working memory span.

Figures

Figure 16: Proportion correct recognition in Experiment 4 as a function of age, accent on target contrast set, and accent on other contrast set.

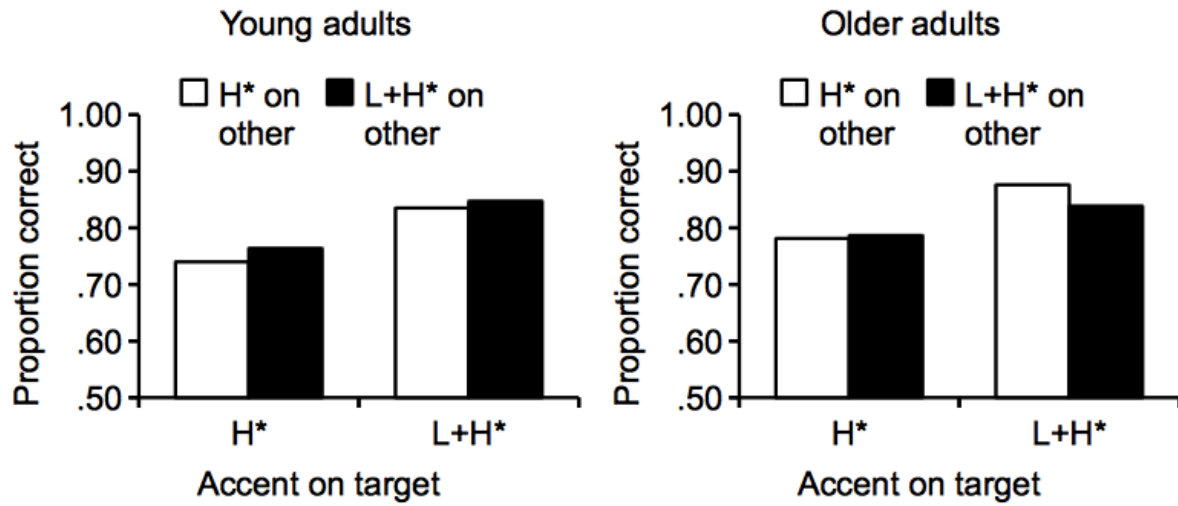


Figure 17: Fixed effect estimates for multi-level logit model of recognition accuracy in Experiment 4 (N = 7680, log-likelihood = -3523).

Fixed effect	β	SE	Wald z	p
Intercept	1.66	0.08	20.63	<.0001
L+H* accent on this word	0.58	0.06	9.40	<.0001
L+H* accent on other word	-0.02	0.06	-0.31	.76
Age	0.15	0.16	0.91	.36
L+H* accent on this word x L+H* on other word	-0.19	0.12	-1.58	.12
Age x L+H* accent on this word	-0.06	0.12	-0.53	.60
Age x L+H* accent on other word	-0.26	0.12	-2.12	<.05
Age x L+H* accent on this word x L+H* on other word	-0.33	0.25	-1.32	.19

Note. SE = standard error.

Figure 18: Summary of random participant and item effects and correlations in model of recognition accuracy in Experiment 4.

Random effect	s^2	Correlation with random intercept
Participant		
Intercept	0.43	
Item		
Intercept	0.05	
Age	0.30	-.66

CHAPTER 6

EXPERIMENT 5

Experiment 5 sought to tease apart the limited resources and linguistic knowledge accounts of older adults' other-accent penalty. The limited resources hypothesis proposes that older adults' selectivity for contrastively accented referents occurs because the process of encoding a focus semantic value is time-consuming and does not allow older adults with the time or resources to encode the rest of the discourse. This hypothesis is consistent with the fact that encoding a focus semantic value required additional reading time in Experiment 2. This hypothesis predicts that young adults with low scores on complex span tasks should show a similar pattern as the older adults in Experiment 4.

By contrast, the linguistic knowledge hypothesis—that older adults' selectivity is driven by greater familiarity with the distribution of contrastive accents—makes no prediction that young adults with low span scores should behave like older adults in this task. If anything, *high* span individuals should be most apt to show the other-accent penalty, since these individuals typically make greater use of constraints in online language processing.

Thus, in Experiment 5⁷, I tested how complex span scores related to the other-accent penalty in young adults. More broadly, this experiment provided a further test of how salient alternatives are processed by a range of listeners and whether additional online processing

⁷ This chapter was adapted from previously published material. Copyright © 2012 by the American Psychological Association. Adapted with permission. The official citation that should be used in referencing this material is 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. doi: 10.1037/a0024138. The use of this information does not imply endorsement by the publisher. I thank my co-authors, Duane Watson and Aaron Benjamin, for their assistance with the design of the research and the preparation of the manuscript.

resources are required to realize the benefits of prominence in discourse.

Method

Participants

56 students at the University of Illinois participated in partial fulfillment of a course requirement or for a cash honorarium.

Materials

The materials for the prosody and memory task were the same as in Experiment 4.

Materials for the reading and listening span tasks were taken from Stine and Hindman (1994) and comprised sentences that defined common English nouns, as in (11) and (12). Half of the statements were true, such as (11), and half were false, such as (12). The two tasks used different sentences. The spoken sentences used in the listening span task were recorded by a different female research assistant.

(11) An elected official who manages a state is called a governor.

(12) One animal that is bright orange in color is the zebra.

Procedure

Participants first completed four complex span tasks, followed by the prosody and memory task from Experiment 4. In each of the four complex span tasks, detailed individually below, each trial consisted of a series of stimuli of varying span length (e.g., a trial with 2 stimuli had span length 2). At the end of a trial, participants were asked to recall some aspect of the stimuli by typing their answers.

Following the recommendations of Conway, Kane, Bunting, Hambrick, Wilhelm, and

Engle (2005), all participants were presented with two trials at all span lengths in a random order. Conway et al. have argued that this format offers multiple advantages over a traditional format in which participants attempt span lengths in ascending order until the participant is unable to recall all the stimuli. First, performance typically decreases over multiple memory tests due to proactive interference from previous tests. Presenting spans in order of ascending length confounds span length with amount of proactive interference, and differences in span score might reflect differences in vulnerability to proactive interference rather than working memory *per se* (Lustig, May, & Hasher, 2001). Second, participants may succeed or fail at a particular span length for reasons unrelated to working memory (e.g., the idiosyncratic memorability of particular words). Presenting all spans to all participants maximizes the amount of information attained from each participant.

Listening span. Participants listened to recorded sentences and then pressed one of two keys to indicate whether the statement was *true* or *false*. Participants were allowed up to 2000 ms after the end of the sentence to make the judgment. The targets were the last words from each sentence. The span length ranged from 2 to 7.

Reading span. Participants read aloud a sentence and pressed one of two keys to indicate whether the statement was *true* or *false*. Participants were allowed up to 7000 ms to read the sentence and make the judgment. The targets to be remembered were the last words from each sentence, such as *governor* in (11). The span length ranged from 2 to 7.

Alphabet span. Following Waters and Caplan (2003), participants read aloud single words and then recalled them in alphabetical order. Each word was displayed visually for 1000 ms. The span length varied from 2 to 7.

Subtract 2 span. Also following Waters and Caplan (2003), participants read aloud

digits from 2 to 9 and then recalled them in order while subtracting 2 from each number (e.g., 3 was to be recalled as 1). Each digit was displayed visually for 1000 ms. The span length varied from 2 to 8.

Scoring

Scores on each of the four complex span tasks were computed as followed. Trials in which the participant remembered all of the items were scored as 1 point. Trials in which the participant remembered some but not all of the items were scored as the proportion of items correctly recalled; for instance, a participant who remembered 3 items from a span 4 trial would receive a score of 0.75. In a comparison of multiple scoring methods, Conway et al. (2005) found this method (termed *partial-credit unit scoring*) to produce the least skewed, most normal distribution of scores.

Finally, the mean of each participant's scores on the four complex span tasks was taken to create an aggregate working memory measure. Aggregating over multiple tasks has the advantage of reducing variance in scores due to task-specific factors (e.g., familiarity with the alphabet) unrelated to the construct of interest (Waters & Caplan, 2003).

Results

Mean performance on the discourse memory task is displayed in Figure 19 and on the complex span tasks in Figure 20.

Average working memory score was entered as a continuous predictor at the subject level, as were the interactions of working memory with the pitch accenting variables. Including working memory score as a continuous predictor, rather than classifying participants into *high* and *low* groups, increases statistical power and accurately reflects the fact that span scores vary continuously rather than categorically in the population (Conway et al., 2005).

The fit of the model was improved by a random slope of target accent by item, $\chi^2_{(2)} = 6.81$, $p < .05$, and further improved by a random slope of other-accent penalty by item, $\chi^2_{(2)} = 9.55$, $p < .05$. No other random slopes approached significance.

Parameter estimates for the final model are displayed in Figures 21 and 22. The odds of correct recognition for contrast sets receiving contrastive accents ($M = 85\%$) were 1.65 times greater (95% CI: [1.38, 1.98]) than for sets receiving presentational accents ($M = 79\%$), replicating the accent boost observed in Experiment 4. Across all participants, the accent on the other critical word did not reliably affect recognition, consistent with Experiment 4, in which young adults as a whole did not display the other-accent penalty.

The effects of a low working memory span mirrored those of age in Experiment 4. Working memory span did not affect the size of the accent boost, Wald $z = -0.01$, $p = .99$, but it did modulate the size of the other-accent penalty, Wald $z = 2.51$, $p < .05$. The size of the other-accent penalty on recognition is captured in the odds ratio between a contrastive accent on the other critical word and a presentational accent on that word. This ratio was 0.84 times smaller (95% CI: [0.74, 0.96]) for every one-point increase in mean working memory score in the model. That is, the other accent penalty was smallest for participants with high working memory scores and largest for participants with low scores.

Working memory score also had a main effect on performance. A one-point increase in mean working memory score translated to a 1.47 times increase in the odds of correct recognition (95% CI: [1.18, 1.84]).

Discussion

In Experiment 5, I pitted two accounts of the age differences in Experiment 4 against each other by testing young adults who varied in their scores on complex span tasks.

The results provided a conceptual replication of the effects in Experiment 4. Young participants in Experiment 5 with *lower* working memory ability resembled the older participants in Experiment 4: they showed an equivalent benefit from a contrastive accent on the target item, but displayed the other-accent penalty to memory when a different part of the discourse received a contrastive accent.

These results support a processing resources account of the age effects. The most important information in a discourse may always be processed and encoded even when online processing resources such as working memory are limited. However, online resources may restrict the ability to encode less important information along with the focus semantic value. Consequently, only those participants with greater processing resources—the higher-span young adults—do not show an other-accent penalty.

The lower-span young adults in Experiment 4 did differ from the older adults in Experiment 3 in one respect. While older adults did not differ from young adults in their mean level of performance, lower-span young adults had lower overall memory than higher-span young adults. One possibility is that the relationship between working memory score and discourse memory reflects variance shared with a more basic construct such as motivation or arousal, which could affect performance on both the working memory and discourse tasks.

More broadly, the results of Experiment 5 provide further evidence that encoding a focus semantic value is an effortful, time-consuming process. Participants with more limited ability to complete processing tasks were less successful in encoding the rest of the discourse after they had encoded a focus semantic value.

Figures

Figure 19: Proportion correct recognition in Experiment 5 as a function of accent on target contrast set and accent on other contrast set.

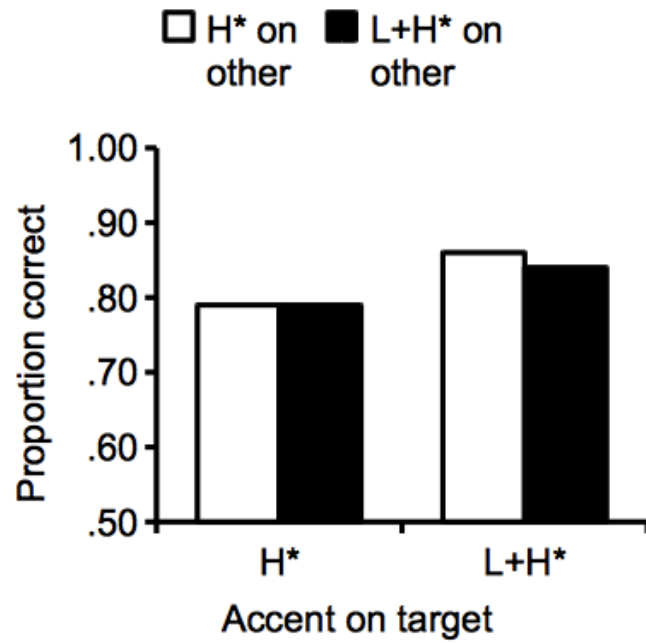


Figure 20: Mean score on span tasks in Experiment 5. The scoring procedure is described in the Method of Experiment 5. Error bars indicate one standard deviation in each direction.

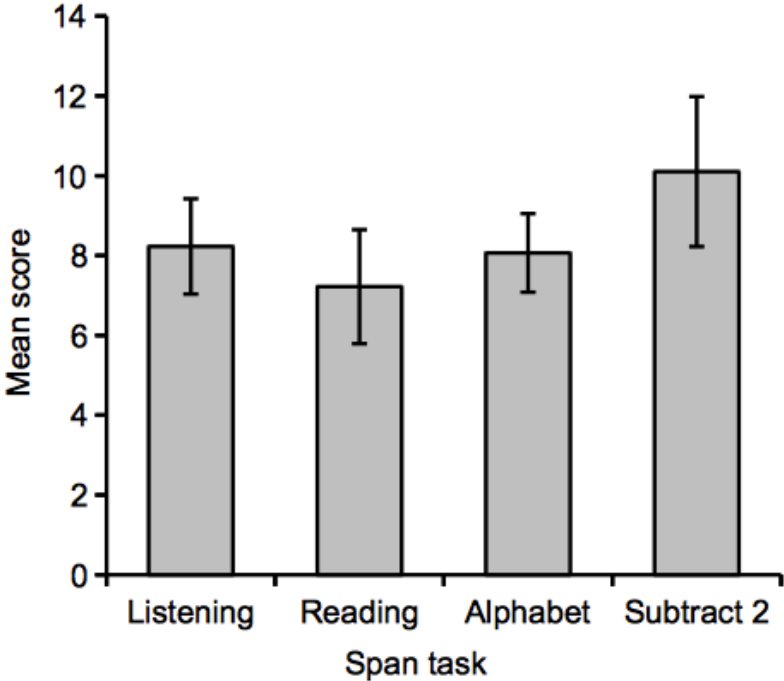


Figure 21: Fixed effect estimates for multi-level logit model of recognition accuracy in Experiment 5 (N = 4480, log-likelihood = -1908).

Fixed effect	β	SE	Wald z	p
Intercept	1.85	0.13	13.99	<.0001
L+H* accent on this word	0.50	0.09	5.41	<.0001
L+H* accent on other word	-0.01	0.09	-0.09	.94
Working memory (WM) score	0.39	0.11	3.37	<.0001
L+H* accent on this word x L+H* on other word	-0.15	0.17	-1.18	.38
WM x L+H* accent on this word	> -0.01	0.07	-0.03	.99
WM x L+H* accent on other word	-0.17	0.07	-2.46	<.05
WM x L+H* accent on this word x L+H* on other word	-0.11	0.14	0.76	.40

Note. SE = standard error.

Figure 22: Summary of random participant and item effects and correlations in model of recognition accuracy in Experiment 5.

Random effect	s^2	Correlations	
		1	2
Participant			
Intercept	0.73		
Item			
1. Intercept	0.16	—	
2. L+H* on this word	0.10	.87	—
3. L+H* on other word	0.11	>.99	-.86

CHAPTER 7

GENERAL DISCUSSION

In five experiments, I investigated how salient alternatives in a discourse affect both online processing and later memory. Experiment 1 generalized past findings from contrastive (L+H*) pitch accents to font emphasis in text: Emphasizing a word in a discourse helped readers reject memory probes referring to a salient alternative to the emphasized word but not memory probes referring to an unmentioned item, suggesting the effect of emphasis was to lead readers to encode something about the salient alternative, or *focus semantic value*. Experiment 2 then provided evidence that this mnemonic benefit was contingent upon readers devoting extra reading time to the emphasized words. This benefit was not replicated in Experiment 3; however, Experiments 4 and 5 provided other evidence in favor of a relation of online processing to later memory. In Experiments 4 and 5, populations with more restricted online processing ability—younger adults with lower scores on complex span tasks and older adults—showed impaired memory for the *rest* of the discourse in older adults when they had occasion to encode a focus semantic value. Finally, Experiment 3 provided evidence that readers encode a relatively narrow set of alternatives in response to the font emphasis: font emphasis only conferred a benefit in rejecting only a particularly salient alternative to the true proposition, and not a referent that was merely mentioned in the discourse.

Emphasized Text and Memory

Experiments 1 through 3 tested the effect of prominence in written discourse. 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 false statement

regarding a salient alternative. This pattern suggests that, at least given a salient alternative present in the discourse, the memory benefit of font emphasis stems from encoding of a focus semantic value, or set of alternatives to a true proposition (Rooth, 1992).

A similar benefit to memory has been observed in response to contrastive (L+H*) pitch accents (Fraundorf et al., 2010, and the present Experiments 4 and 5). Because evoking contrast has been argued to be the idiosyncratic effect of the L+H* intonational contour (Gussenhoven, 1983; Pierrehumbert & Hirschberg, 1990), it was possible that these memory benefits would require an explicit presentation of the L+H* pitch contour. The present results, however, suggest that the effect is more general, because it also obtains 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., Fodor, 1998; Breen & Clifton, 2010), they may have generated the L+H* accent in their implicit prosody when they encountered the emphasized text. Another explanation is based on recent proposals (Calhoun, 2009) that *any* linguistic element that is more prominent than expected can promote the consideration of salient alternatives. In this account, the prominence of a word with font emphasis could lead to consideration of a salient alternative even without implicitly evoking a contrastive pitch accent. It will be necessary for future research to directly compare these hypotheses, perhaps by attempting to eliminate implicit prosody through articulatory suppression techniques (for review, see Besner, 1987). Nevertheless, the present experiments show that the representation of salient alternatives in discourse does not depend on an overt manipulation of contrastive prosody.

The results of the present experiments are also noteworthy for demonstrating clear

memory benefits from font emphasis. Prior results have been somewhat mixed as to whether emphasized text actually improves learning from a text (for review, see, Hartley, Bartlett, & 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 reject only certain kinds of false information. Consistent with this, Golding and Fowler (1992) found that underlined words—in concert with several other manipulations designed to make important details more prominent—only benefited learning of a text when readers were told that they would be tested on specific details rather than the overall gist.

Perceptual Characteristics of Prominent Information

Although the present experiments suggest that prominence leads readers to encode a focus semantic value, font manipulations and prosodic pitch accents both also alter the perceptual properties of words. One possibility is that these changes simply made the prominent 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 be slower or more difficult to read, and 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 prominence was not to modulate the representation of the target word, but of the salient alternative. If the prominence of *British* led to enhanced representation of *British* itself, that should have facilitated rejection of all false lures, all of which would be inconsistent with remembering *British* (Brainerd et al., 2006). Instead, emphasizing *British* only enhanced

rejections of the contrast lure and not unmentioned lures. This implies the effect of the font manipulation did not lie in the representation of *British* itself. Furthermore, 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 likely that the perceptual characteristics of prominent words play a role in how they are processed. Perceptual distinctiveness or difficulty may be a cue that additional processing is needed. In the presence of a clear contrast set, however, it appears the effect of prominence is not to globally enhance memory but to engender a contrastive reading. 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.

What Constitutes a Focus Semantic Value?

The present experiments also tested how wide a set of salient alternatives comprises a focus semantic value. Rooth (1992) has proposed that the set of alternatives that comprise a focus semantic value are contextually constrained, but it has been unclear whether the context supports a relatively broad band of alternatives—all items in the same semantic category, for instance, or all those mentioned in the discourse—or only a narrow set of the most likely alternatives. The present experiments supported the latter hypothesis. In Experiments 1 and 2, font emphasis facilitated later rejections of a salient alternative in the discourse but not of an item from the same semantic category but unmentioned in the discourse. In Experiment 3, an even stricter test, font emphasis did not facilitate rejections of an item that was mentioned in the discourse, but in a context that made it a less plausible alternative. These results imply that comprehenders consider and encode only a narrow set of alternatives in a focus semantic value.

This result is consistent with results from linguistic other domains indicating that discourses can tightly constrain the number of relevant alternatives. For example, speakers frequently produce seemingly ambiguous referring expressions so long as the alternative referents are physically distant or task-irrelevant and thus outside the domain of reference (Brown-Schmidt & Tanenhaus, 2008). That is, *the green block* can be produced even in the presence of multiple green blocks when discourse constraints have restricted the domain of reference to include just one of the green blocks. Addressees, in turn, interpret these “ambiguous” referring expressions without difficulty, indicating only a restricted set of possible referents was considered in comprehension as well (Brown-Schmidt & Tanenhaus, 2008).

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 mere familiarity or prior mention. In Experiment 3, both the alternative probes and the merely mentioned probes had been mentioned in the same discourse, but only rejections of the alternative probes showed a benefit from the font emphasis. Thus, the locus of the emphasis effect was in its ability to reject alternatives that would likely be in a focus semantic value.

The Role of Contrast in Cognition

Although the present study focused on how representing salient alternatives contributed to long-term, offline memory for a discourse, 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 a noun phrase containing a scalar adjective, such as *the small--*, they preferentially look to an object that has an alternative that differs in size (e.g., a small glass when there is also a large glass present)

rather than a singleton object without a salient alternative (e.g., a small bowl with no large bowl present; Sedivy et al., 1999). This looking preference obtains even if the alternative is no longer visually copresent (Wolter, Gorman, & Tanenhaus, 2011), providing further evidence that salient alternative sets are maintained in memory. Similar results obtain when contrastive (L+H*) pitch accents are placed on color adjectives (Ito & Speer, 2008) or on bare noun phrases that have a cohort competitor present (a *camel* when there is also a *candle*; Watson et al., 2008).

The computation of salient 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).

Online Language Comprehension and Long-Term Memory

A major question of interest in the present study was how the long-term mnemonic benefits of focus semantic values was tied to online processing of prominent information. This relation was assessed in Experiments 2 through 5; three of these four experiments provided evidence suggesting that computing a focus semantic value required additional online processing time. In Experiment 2, the mnemonic benefits of font emphasis increased the more time readers spent on the emphasized information. In Experiments 4 and 5, participants listened to spoken discourses; older adults and some young showed decreased memory for less prominent information when contrastive prosody presumably led them to compute a focus semantic value. This relation suggests that computing the focus semantic value was time-consuming and left participants with less time to encode the rest of the discourse. However, Experiment 3 provided conflicting results; in this experiment, no relation of online reading time to later memory was observed. Given the similarity of Experiments 2 and 3 in materials and procedures, it is unclear

what would account for these discrepant findings, but I note that past investigations of the relation between online reading time and offline discourse comprehension have produced similar mixed results (e.g., Caplan et al., 2011; Christianson & Luke, 2011; Daneman et al., 2007, Reder & Kusbit, 1991; Ward & Sturt, 2007).

More generally, the present study highlights point a need for more consideration of how online language processing time relates to later comprehension. The literature on the allocation of study time suggests several important points that may influence accounts of how reading time relates to language comprehension.

First, the relationship of study or reading time to later understanding can vary. Some theories of language comprehension posit that readers always arrive at a fully specified representation, with that representation requiring more or less time to construct depending on its difficulty and the need for revision. In these accounts, additional reading time is invested in the face of difficulty to maintain a constant level of comprehension (what would be termed a *discrepancy reduction* model in the literature on study time allocation; e.g., Dunlosky & Hertzog, 1998). However, study time can relate to comprehension in other ways. In some cases, additional study time may be applied *without* benefit to comprehension, known as the *labor-in-vain* effect (Nelson & Leonesio, 1998). For example, when participants need to resolve global syntactic ambiguities in order to answer comprehension questions, they read those ambiguities more slowly, yet still have difficulty answering the questions (Swets, Desmet, Clifton, & Ferriera, 2008). In other cases, as in the present Experiment 2 and Caplan et al. (2011), additional reading time may *increase* the depth of processing, with slower reading of key material actually predicting better comprehension. These findings are consistent with the principle that initial ease of processing is poorly predictive of, and may even be negatively

related to, later retention (Schmidt & Bjork, 1992).

Second, the application of study time may be inconsistent or inefficient. While learners *typically* devote more study time to difficult material, they do not *always* do so (Son & Kornell, 2008; Son & Metcalfe, 2000). Moreover, differences in the ability to selectively devote more study time where needed can partially or wholly account for differences in memory performance between individuals (Tullis & Benjamin, 2010) and between age groups (Dunlosky & Connor, 1997). In the present experiment, readers did not always spend additional time generating a relevant contrast, and memory was impaired when they did not. That is, although additional reading time would have facilitated comprehension of the discourse, readers did not always apply this extra time. This pattern is consistent with a good-enough approach to language comprehension in which the comprehension system does not always construct fully specified representations.

Conversely, the present results also speak to the important role 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, longer reading times to emphasized words only benefited rejection of certain kinds of lures: those that the discourse had established as particularly salient alternatives. This result joins 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). And, although repeated words are generally easier to process, this effect can be reversed when the discourse context makes lexical repetition less

felicitous than a pronoun (Ledoux, Gordon, Camblin, & Swaab, 2007).

Discourse Comprehension in Older Adults

Although older adults in Experiment 4 exhibited decreased memory for non-prominent information, as noted above, they showed just as much benefit as young adults from a contrastive accent on a target word. This pattern can be viewed as strategic and adaptive: older adults, given their more restricted processing abilities, wisely chose to focus on the details that the speaker had emphasized, even if it came at the cost of other, less prominent information.

Why are older adults less strategic than young adults in some memory tasks (e.g., Dunlosky & Connor, 1997) but equally strategic in others? Clearly, more research must be done to determine when older adults do and do not succeed in strategic memory encoding. But, one possible moderating variable may be the presence of external cues supporting selectivity. Experiments that have found equivalent selectivity across the life span typically have included cues to the importance of information, such as pitch accents (in the present work) or point values (Castel et al., 2002). It has been proposed that age differences in memory are greater for tasks or materials that do not require controlled or self-initiated processing (e.g., Craik, 1983, 1986). External cues to importance like pitch accents or point values may allow older adults to exhibit selective control of memory by reducing the need to initiate selective processing on one's own. This explanation is also consistent with age differences on acquiring new metacognitive knowledge. Older adults, unlike young adults, often do not appear to learn about the effectiveness of different strategies from experience. However, they do learn about the difficulty of various types of items (e.g., words of different frequency) when the item type can be discerned from the stimulus itself (Tullis & Benjamin, in press). This pattern is consistent with a view in which older adults' metacognitive success depends in part of the available of cues in the

environment.

The discrepant results concerning selectivity also underscore the importance of testing of older adults' memory with naturalistic materials. Pitch accents are a common cue to importance with a discourse. If older adults make use of such frequently occurring cues, then their ability to strategically process a discourse may be greater than thought.

What Underlies Online Processing Resources?

The similarity in this task between young adults who score low on complex span tasks and older adults supports a processing resources account of age differences in prosody use. One question that might be asked is exactly what underlies these differences in resources. Variance in complex span task performance between individuals or across the life span has frequently been attributed to more fundamental cognitive constructs such as processing speed (Salthouse, 1996), executive control (Engle, 2002), linguistic knowledge or skills (MacDonald & Christiansen, 2002), or inhibitory processing (Hasher et al., 1999).

In the present work, I used measures of working memory as a proxy for online processing resources. It is entirely possible that a more basic construct ultimately accounts for the difference between age groups. For instance, the age-related changes in working memory that bear on language processing have been modeled as a function of more fundamental decreases in processing speed (Stine-Morrow et al., 2008).

It is noteworthy that the older adults in this task resembled the *low*-span young adults in spite of having greater vocabulary. In young adults, working memory scores typically correlate positively with vocabulary; this correlation has been interpreted as suggesting that many effects attributed to working memory may instead reflect linguistic experience (MacDonald & Christiansen, 2002). However, the fact that older adults resemble *low*-span young adults in this

task suggests that biological changes across the lifespan may sometimes be a stronger influence than gains in linguistic knowledge (see MacDonald & Christiansen, 2002, for further discussion).

Conclusion

Font emphasis, just like some pitch accents, can lead readers to represent information about salient alternatives in a discourse. Knowledge of these alternatives can benefit later memory by helping to rule out those alternatives as candidates for what actually happened. . Focus semantic values appear to be narrowly defined, with emphasis benefiting rejections of only the most plausible alternatives. Moreover, the representation of salient alternatives in discourse appears relatively pervasive and is not unique to contrastive prosody; it can also be generated by font emphasis in written text.

Three of four experiments suggested that these focus semantic values require time to calculate and are not always generated. Consequently, both young adults with lower complex span scores and older adults, both of whom may not have the online processing resources to encode everything in a discourse, exhibited decreased memory for the rest of the discourse when element is emphasized. In addition, in some cases, representations of a focus semantic value were more likely to be observed when participants spent more time reading the emphasized words, although this relation was not always observed.

These findings support a view of linguistic devices like prominence exert an important influence on what comprehenders encode and remember over the long-term. Moreover, these devices do not appear to exert a uniform influence on long-term memory. Rather, some aspects of a discourse representation are sometimes left unspecified and require an additional investment of time to calculate.

REFERENCES

- Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Hoboken, NJ: John Wiley & Sons.
- 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. doi:10.1016/j.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. doi: 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: Vol. 48. Skill and strategy in memory use* (pp.175-223). London: Academic Press. doi: 10.106/S0079-7421(07)48005-7
- Benjamin, A. S. (2010). Representational explanations of process dissociations in recognition: The DRYAD theory of aging and memory judgments. *Psychological Review*, *117*, 1055-1079. doi:10.1037/a0020810
- 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. doi: 10.1080/14640748708401799
- Birch, S., & Rayner, K. (2010). Effects of syntactic prominence on eye movements during reading. *Memory & Cognition*, *38*, 740-752. doi: 10.3758/MC.38.6.740
- Bock, J. K., & Mazzella, J. E. (1983). Intonational marking of given and new information: Some consequences for comprehension. *Memory & Cognition*, *11*, 64-76. doi: 10.3758/BF03197663
- Brainard, D. H. (1997) The Psychophysics Toolbox, *Spatial Vision*, *10*, 433-436. doi:10.1163/156856897X00357

- Brainerd, C. J., Reyna, V. F., & Estrada, S. (2006). Recollection rejection of false narrative statements. *Memory, 14*, 672-691. doi:10.1080/09658210600648449
- 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. doi: 10.1080/03640210802066816
- 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. doi: 10.1037//a0021837
- Castel, A. D. (2008). The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. In A. S. Benjamin and B. H. Ross (Eds.), *The Psychology of Learning and Motivation: Vol. 48. Skill and strategy in memory use* (pp. 225-270). San Diego, CA: Academic Press. doi: 10.1016/S0079-7421(07)48006-9

- Castel, A. D., Benjamin, A. S., Craik, F. I. M., & Watkins, M. J. (2002). The effects of aging on selectivity and control in short-term recall. *Memory & Cognition*, *30*, 1078-1085. doi: 10.3758/BF03194325
- Christianson, K., & Luke, S. G. (2011). Context strengthens initial misinterpretations of text. *Scientific Studies of Reading*, *15*, 136-166. doi: 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. doi: 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. doi: 10.1177/01466216830070030
- Cohen, G., & Faulkner, D. (1986). Does 'elderspeak' work? The effect of intonation and stress on comprehension and recall of spoken discourse in old age. *Language & Communication*, *6*, 91-98. doi: 10.1016/0271-5309(86)90008-X

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide.

Psychonomic Bulletin & Review, *12*, 769-786. doi: 10.3758/BF03196772

Craik, F. I. M. (1983). On the transfer of information from temporary to permanent memory.

Philosophical Transactions of the Royal Society of London, Series B, *302*, 341-359.

Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix & H.

Hagendorf (Eds.), *Human memory and cognitive capabilities, mechanisms and performances* (pp. 409-422). Amsterdam: Elsevier.

Daneman, M., Lennertz, T., & Hannon, B. (2007). Shallow semantic processing of text:

Evidence from eye movements. *Language and Cognitive Processes*, *22*, 83-105. doi:

10.1080/01690960500372725

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A met-

analysis. *Psychonomic Bulletin & Review*, *3*, 422-433. doi: 10.3758/BF03214546

Davis, T., & Love, B. C. (2010). Memory for category information is idealized through contrast

with competing options. *Psychological Science*, *21*, 234-242. doi:

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. doi:10.1016/j.cognition.2010.09.012
- Dixon, R. A., Hultsch, D. F., Simon, E. W., & von Eye, A. (1984). Verbal ability and text structure effects on adult age differences in text recall. *Journal of Verbal Learning and Verbal Behavior, 21*, 569-578. doi: 10.1016/S0022-5371(84)90355-4
- 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. doi: 10.3758/BF03211311
- Dunlosky J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 249-275). Mahwah, NJ: Erlbaum.
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19-23. doi: 10.1111/1467-8721.00160

Farmer, T. A., Christiansen, M. H., & Kemtes, K. A. (2005). Sentence processing in context: The impact of experience on individual differences. In *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp 642-647). Mahwah, NJ: Lawrence Erlbaum Associates.

Ferreira, F., & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348-368. doi: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. doi: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. doi: 10.1518/001872006776412199

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

Folstein, M. F., Robins, L. N., & Helzer, J. E. (1983). The mini-mental state examination. *Archives of General Psychiatry*, 40, 812.

Font (2011). *Oxford English Dictionary*. Retrieved August 11, 2011, from <http://www.oed.com.proxy2.library.illinois.edu/view/Entry/73956> .

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. doi: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. doi: 10.1037/a0024138

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. doi: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. doi: 10.1016/0361-476X(92)90052-Z

Graesser, A. C., Mills, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, *48*, 163-189. doi: 10.1146/annurev.psych.48.1.63

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

- 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. doi: 10.1017/S0022226700016613
- Hartley, J. T. (1993). Aging and prose memory: Tests of the resource-deficit hypothesis. *Psychology and Aging*, 8, 538-551. doi: 10.1037/0882-7974.8.4.538
- Hartley, J., Bartlett, S., & Branthwaite, A. (1980). Underlining can make a difference—sometimes. *The Journal of Educational Research*, 73, 218-224.
- 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. doi: 10.1037/0022-0663.90.3.414
- Hasher, L., Zacks, R. T., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriat (Eds.), *Attention and performance* (Vol. 17, pp. 653-675). Cambridge, MA: MIT Press.
- 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. doi: 10.1016/S0022-5371(74)80003-4

- Healey, M. K., Campbell, K. L., & Hasher, L. (2008). Cognitive aging and increased distractibility: costs and potential benefits. In W. S. Sossin, J.-C. Lacaille, V. F. Castellucci, & S. Belleville (Eds.), *Progress in brain research* (Vol. 169, pp. 353-363). doi: 10.1016/S0079-6123(07)00022-2
- 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. doi: 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. doi: 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. doi: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. doi:10.1016/j.jml.2007.11.007

- Just, M. A., & Carpenter, P. A. (1983). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*, 228-238. doi:10.1037/0096-3445.111.2.228
- Kemper, S., & Harden, T. (1999). Experimentally disentangling what's beneficial about elderspeak from what's not. *Psychology and Aging*, *14*, 656-670. doi: 10.1037/0882-7974.14.4.656
- Kjelgaard, M. M., Titone, D. A., & Wingfield, A. (1999). The influence of prosodic structure on the interpretation of temporary syntactic ambiguity by young and elderly listeners. *Experimental Aging Research*, *25*, 187-207. doi: 10.1080/036107399243986
- Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, *14*, 219-224. doi: 10.3758/BF03194055
- Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: phonetics, phonology, and sound change*. New York: Walter de Gruyter.
- Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge: Cambridge University Press.
- 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. doi: 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. doi: 10.1006/ceps.1995.1003
- Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General, 130*, 199-207. doi: 10.1037//0096-3445.130.2.199
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comments on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*, 35-54. doi:10.1037//033-295X.109.1.35
- 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. doi: 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. doi: 10.1080/09658211.2010.530271

- McAteer, E. (1992). Typeface emphasis and information focus in written language. *Applied Cognitive Psychology*, 6, 345-359. doi: 10.1002/acp.2350060406
- McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, 114, 1-16. doi: 10.1037/0096-3445.114.1.1
- Nelson, T. O., & Leonesio R. J. (1988). Allocation of self-paced study time and the “labor-in-vain effect.” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 676-686. doi: 10.1037/0278-7393.14.4.676
- Orbelo, D. M., Grim, M. A., Talbott, R. E., & Ross, E. D. (2005). Impaired comprehension of affective prosody in elderly subjects is not predicted by age-related hearing loss or age-related cognitive decline. *Journal of Geriatric Psychiatry and Neurology*, 18, 25-32. doi: 10.1177/0891988704272214
- Orbelo, D. M., Testa, J. A., & Ross, E. D. (2003). Age-related impairments in comprehending affective prosody with comparison to brain-damaged subjects. *Journal of Geriatric Psychiatry and Neurology*, 16, 44-52. doi: 10.1177/0891988702250565
- Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, 17, 299-320. doi: 10.1037//0882-7974.17.2.299

- Pearlmutter, N. J., & MacDonald, M. C. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language*, *34*, 521–542. doi:10.1006/jmla.1995.1024
- Pelli, D. G. (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437-442. doi: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.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372-422. doi: 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. doi: 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. doi: 10.3758/PBR.16.1.1

- Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, *1*, 75-116. doi: 10.1007/BF02342617
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*, 403-428. doi: 10.1037/0033-295X.103.3.403
- Salthouse, T. A. (2004). What and when of cognitive aging. *Current Directions in Psychological Science*, *13*, 140–144. doi:10.1111/j.0963– 7214.2004.00293.x
- 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. doi: 10.3758/MC.37.6.880
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: not noticing the evidence. *TRENDS in Cognitive Sciences*, *6*, 382-386. doi: 10.1016/S1364-6613(02)01958-7
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, *3*, 207-217. doi: 10.1111/j.1467-9280.1992.tb00029.x
- Sedivy, J. C. (2002). Invoking discourse-based contrast sets and resolving syntactic ambiguities. *Journal of Memory and Language*, *46*, 341-370. doi: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. doi: 10.1016/S0010-0277(99)00025-6
- Selkirk, E. (2002). Contrastive *FOCUS* vs. presentational *focus*: Prosodic evidence from right node raising in English. *Speech prosody*, Aix-en-Provence, 643-646.
- Shipley, W. C. (1940). A self-administered scale for measuring intellectual impairment and deterioration. *Journal of Psychology*, *9*, 371-377. doi: 10.1080/00223980.1940.9917704
- Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (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., & Kornell, N. (2008). Research on the allocation of study time: Key studies from 1890 to the present (and beyond). In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 333-351). New York: Psychology Press.
- 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. doi:10.1037//0278-7393.26.1.204
- Steedman, M. (2000). Information structure and the syntax-phonology interface. *Linguistic Inquiry*, *31*, 649-689. doi: 10.1162/002438900554505

- Stine, E. A. L., & Hindman, J. (1994). Age differences in reading time allocation for propositionally dense sentences. *Aging and Cognition, 1*, 2-16. doi: 10.1080/09289919408251446
- Stine, E. A. L., & Wingfield, A. (1987). Process and strategy in memory for speech among younger and older adults. *Psychology and Aging, 2*, 272-279. doi: 10.1037/0882-7974.2.3.272
- Stine, E. A. L., & Wingfield, A. (1988). Memorability functions as an indicator of qualitative age differences in text recall. *Psychology and Aging, 3*, 179-183. doi: 10.1037/0882-7974.3.2.179
- Stine-Morrow, E. A. L., Miller, L. M. S., Gagne, D. D., & Hertzog, C. (2008). Self-regulated reading in adulthood. *Psychology and Aging, 23*, 131-153. doi: 10.1037/0882-7974.23.1.131
- Swets, B., Desmet, T., Clifton, C., Jr., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition, 36*, 201-216. doi: 10.3758/MC.36.1.201

- 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. doi: 10.1037/0278-7393.25.4.1024
- Titone, D. A., Koh, C. K., Kjelgaard, M. M., Bruce, S., Speer, S. R., & Wingfield, A. (2006). Age-related impairments in the revision of syntactic misanalyses: Effects of prosody. *Language and Speech*, *49*, 75-99. doi: 10.1177/002383096060490010501
- Titone, D. A., Prentice, K. J., & Wingfield, A. (2000). Resource allocation during spoken discourse processing: Effects of age and passage difficulty as revealed by self-paced listening. *Memory & Cognition*, *28*, 1029-1040. doi: 10.1177/00238309060490010501
- Tullis, J. G., & Benjamin, A. S. (2010). On the effectiveness of self-paced learning. *Journal of Memory and Language*, *64*, 109-118. doi:10.1016/j.jml.2010.11.002
- Tullis, J. G., & Benjamin, A. S. (in press). The effectiveness of updating metacognitive knowledge in the elderly: Evidence from metamnemonic judgments of word frequency. *Psychology and Aging*. doi: 10.1037/a0025838
- Ward, P., & Sturt, P. (2007). Linguistic focus and memory: An eye movement study. *Memory & Cognition*, *35*, 73-86. doi: 10.3758/BF03195944

- Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers*, *35*, 550-564. doi: 10.3758/BF03195534
- Watson, D., Gunlogson, C., & Tanenhaus, M. (2008). Interpreting pitch accents in on-line comprehension: H* vs L+H*. *Cognitive Science*, *32*, 1232-1244. doi: 10.1080/03640210802138755
- Wingfield, A., Lindfield, K. C., & Goodglass, H. (2000). Effects of age and hearing sensitivity on the use of prosodic information in spoken word recognition. *Journal of Speech, Language, and Hearing Research*, *43*, 915-925.
- 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. doi: 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. doi: 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. doi:10.1006/jmla.2002.2864

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 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.
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 Valentine'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

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 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 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

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).
- 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 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).