Reduction in Prosodic Prominence Predicts Speakers' Recall: Implications for Theories of Prosody

Scott H. Fraundorf

University of Pittsburgh

Duane G. Watson and Aaron S. Benjamin

University of Illinois at Urbana-Champaign

Scott H. Fraundorf (corresponding author), University of Pittsburgh, 608 Learning Research and Development Center, Pittsburgh, PA 15260. Telephone: +1 (412) 624-7029. E-mail: scottfraundorf@gmail.com

Duane G. Watson, University of Illinois at Urbana-Champaign, 603 E Daniel St., Champaign, IL, 61820. Telephone: +1 (217) 333-0280. E-mail: dgwatson@illinois.edu

Aaron S. Benjamin, University of Illinois at Urbana-Champaign, 603 E Daniel St., Champaign, IL, 61820. Telephone: +1 (217) 333-6822. E-mail: asbenjam@illinois.edu

Acknowledgement

This work was supported by National Science Foundation under Graduate Research Fellowship 2007053221; and the National Institutes of Health under Grant T32-HD055272, Grant R01 DC008774, and Grant R01 AG02626.

8,384 words

Abstract

Repeated words are often reduced in prosodic prominence, but the underlying mechanisms remain unclear. The present study contrasted two theories: does prosodic reduction reflect the choice of a particular linguistic form, or does ease of retrieval within the language production system lead to facilitated, less prominent productions? One test of facilitation-based theories is suggested by findings on human memory: Whether a second presentation of an item benefits later memory is predicted by the item's availability at the time of the second presentation. If prosodic reduction partially reflects facilitated retrieval, it should predict later memory. One naïve participant described to another participant routes on a map. Critical items were mentioned twice. Following the map task, the speaker attempted written recall of the mentioned items. As expected, acoustic intensity of the second mentions predicted later recall in the same way that difficulty of retrieval has in other tasks. This pattern suggests that one source of prosodic reduction is facilitation within the language production system.

Keywords: prosody; prominence; recall; discourse; language production

Words in running speech vary in their prosodic prominence, or the degree to which they perceptually stand out from surrounding words. Information that is repeated or predictable is typically reduced in prominence, but the mechanisms that underlie this change are unclear. One theory is that prosodic reduction is a grammatical choice by speakers to convey information about discourse structure (Gussenhoven, 1983; Pierrehumbert, 1980; Pierrehumbert & Hirschberg, 1990; Schwarzschild, 1999; Selkirk, 1995, 2002; Steedman, 2000). Another is that reduction is partly the byproduct of the availability of information to a speaker's language production system (Arnold & Watson, in press; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Kahn & Arnold, 2012).

One way to test these accounts is to determine whether speakers' prosodic prominence predicts their later memory for a discourse. Whether a second encounter with an item increases the probability of remembering it later depends on how difficult the item is to retrieve at the time it is re-encountered (e.g., Benjamin & Tullis, 2010; R. Bjork & Allen, 1970; R. Bjork & E. Bjork, 1992; Matzen & Federmeier, 2010). If prosodic prominence partially reflects the difficulty of retrieving the relevant representations, then we might expect that the degree of prosodic prominence, too, should predict the probability of later recall. If, on the other hand, the decision to produce a word with prominence is a categorical choice, we would not expect a relation between relative prominence and the probability of later recall.

In the present study, we test this prediction by investigating whether prosodic prominence on a word during a communicative task predicts the speaker's later explicit memory for what was discussed. We further discuss how different acoustic correlates of prominence may be differentially associated with different cognitive processes, supporting a view in which prosodic

prominence may stem from multiple sources.

Explaining Variation in Prominence

The prominence of a word is reliably correlated with multiple acoustic and segmental characteristics, including duration, intensity, and fundamental frequency (F0) as well as other measures such as vowel quality (for review, Wagner & Watson, 2010). Early work on prosodic prominence found that the second time a word is mentioned, it is, on average, reduced in intensity, duration, and F0 (Fowler & Housum, 1987). More recently, it has been proposed that words are reduced to the extent they are generally predictable (Aylett & Turk, 2004; Jurafsky, Bell, Gregory, & Raymond, 2001), with the degree of reduction being a continuous function of increasing predictability (Bell et al., 2009).

However, many theories do not specify how predictability and repetition effects are cognitively implemented. One way reduction could be implemented is that speakers deliberately choose between reduced or nonreduced linguistic forms; Kahn and Arnold (2012) term such models *trigger models*. For example, in many linguistic accounts of prominence in English, speakers mark new or contrastive information with *pitch accents*, phonological constructs that increase the prominence of particular words (e.g., Gussenhoven, 1983; Pierrehumbert, 1980; Pierrehumbert & Hirschberg, 1990; Schwarzschild, 1999; Selkirk, 1995, 2002; Steedman, 2000). Linguistic rules (e.g., Pierrehumbert, 1980; Selkirk, 1995, 2002; Steedman, 2000) determine when speakers choose accented forms of words or when accented forms are licensed (or when deaccenting is licensed; Schwarzschild, 1999). Although termed *pitch accents*, such prominent forms may be realized with a variety of acoustic cues beyond F0 (Breen, Fedorenko, Gibson, & Wagner, 2010; Wagner & Watson, 2010); indeed, some recent studies (Breen et al., 2010; Cole,

Mo, Hasegawa-Johnson, 2010; Kochanski, Grabe, Coleman, & Rosner, 2005) have found that acoustic duration or intensity may be better predictors of perceived prominence than is F0.

However, reduction in prominence does not necessarily require the choice of a particular linguistic form. In *processing facilitation* theories (e.g., Arnold & Watson, in press; Kahn & Arnold, 2012), prosodic reduction partially reflects the state of speaker-internal production processes, such as how easily particular discourse or lexical representations can be retrieved. Most theories of language production posit multiple levels of representation and processing in language production (for a review, see Griffin & Ferreira, 2006), including both the conceptualization of the discourse and the selection of lexical forms. Kahn and Arnold (2012) proposed that facilitating retrieval at any of these levels of representation should lead to prosodic reduction, and that facilitating retrieval at multiple levels to even greater reduction. For instance, a word that has been recently produced or can be predicted from context is more available and can be more easily retrieved and articulated, resulting in rapid, less intense articulation. By contrast, words that are new and not predictable from context are less easy to plan and articulate, requiring a more effortful, prominent production. Similarly, expressions that refer to predictable or to already mentioned referents are easier to plan and thus more reduced than expressions referring to new, unpredictable referents. Processing facilitation theories do not preclude that *some* variation in prominence might stem from speakers' deliberate choice of particular linguistic forms, but they do claim that at least some aspects of prosodic prominence are driven by the availability of representations in a speaker's internal language production system.

Some evidence suggests that speaker-internal production processes do contribute to prominence. Speakers reduce the prominence of the second mention of a recently spoken word

even if they are speaking to a new addressee who has not heard the first mention (Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon, & Newlands, 2000; Kahn & Arnold, in press; Vajrabhaya & Kapatsinski, 2011). Because the word is new to the current conversation, this pattern implies that such reduction is not being driven by the givenness or predictability in conversation. A plausible explanation is that speakers reduce the second mention of the word in these instances because the previous mention of the word made it easier to retrieve and facilitated speakers' own internal production processes. Indeed, prosodic prominence is also reduced by other manipulations expected to increase the availability of representations relevant to language production. For example, presenting speakers with a prime that activates both linguistic and non-linguistic representations results in greater reduction than a prime that activates only non-linguistic representations (Kahn & Arnold, 2012). And when repetition and predictability are experimentally pitted against each other such that repetition is now unexpected, repeated references are still reduced in prominence even though such references are now the less predictable ones (Lam & Watson, 2010). One explanation for this result is that the prior use of the lexical form facilitated its subsequent retrieval by speakers despite being unexpected in the discourse.

Online Processing and Future Memory

To date, much of the experimental evidence for facilitation-based effects on prosodic prominence has come from studies, such as those reviewed above, in which discourse history or context are experimentally manipulated prior to the target utterance. Although these studies have provided important evidence about the mechanisms underlying prosodic prominence, one potential caveat is that some of these experimental manipulations disrupted the natural

intercorrelations of potential influences on prominence. For instance, repetition and predictability are typically positively correlated (that is, speakers are more likely to refer to previously mentioned entities than to new ones; Arnold, 1998), so an experiment in which they become negatively correlated presents an unnatural situation that does not conform to the statistical contingencies of the natural world. The patterns of prominence observed in these experiments might thus reflect participants' adaptation to the artificial experimental task rather than the mechanisms underlying prosodic prominence in natural speech. Indeed, which variables influence prosodic prominence have been observed to change over the course of some experiments as participants adjust to the experimental task (e.g., Lam & Watson, 2014), making it unclear whether effects observed in past experimental studies are representative of natural speech. (For further discussion, see Lam & Watson, 2014.)

The case for facilitation-based influences accounts of prominence would be strengthened by further evidence from paradigms that do not require disrupting natural contingencies. One such method is to examine speakers' *later* memory for the discourse, which requires no manipulation to intervene prior to the target utterance. Individuals naturally forget some items and remember others, and this variability in memory has frequently been linked to differences in retrieval and processing at the time at which individuals initially encounter the to-be-remembered items. Thus, whether speakers do or do not later remember particular referents in a discourse can provide evidence about those referents' availability at the time the speakers referred to them.

This method may be especially efficacious for examining repeated mention because repetition has been observed to have a strong influence on later recall (for review, Crowder,

1976). Specifically, theories in the human memory literature have proposed that repeated presentations or retrieval attempts potentiate later memory to the extent that the items are not already easy to retrieve at the time of their second presentation (Benjamin & Tullis, 2010; R. Bjork & Allen, 1970; R. Bjork & E. Bjork, 1992; Tullis, Benjamin, & Ross, in press). That is, a second presentation of *vase* when *vase* is already readily available confers little additional benefit on memory, but processing the second *vase* when the first *vase* is harder to retrieve constitutes effective *retrieval practice* (Karpicke & Roediger, 2008) that facilitates another, downstream retrieval attempt. Thus, eventual memory is better when the task intervening between presentations is more difficult (R. Bjork & Allen, 1970), when a second presentation comes only after other intervening items (Benjamin & Tullis, 2010; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Melton, 1967), when fewer retrieval cues are provided at each successive practice trial (Finley, Benjamin, Hays, R. Bjork, & Kornell, 2011), or when the second encounter with an item is a retrieval from memory rather than an overt re-presentation (Karpicke & Roediger, 2008) because all of these situations provide a more difficult form of practice that is more like the demands of a later query of memory. This benefit of effort retrieval on later memory holds even when the to-be-remembered items are pictures presented with their names (Wheeler & Roediger, 1992) or when participants are not informed about the final test (Carpenter & DeLosh, 1996), as in the present study. In a study most analogous to the one presented here, Matzen and Federmeier (2010) recorded electroencephalographic event-related potentials (ERPs) while participants were presented with a list of words, some of which were repeated, and then took a recognition memory test. The best predictor of final memory was the ERP response to the second presentation: Items that were *most* likely to be remembered showed

the *least* reduction of the N400 component, which has been argued to index semantic processing (for review, see Kutas & Federmeier, 2011). That is, more effortful, less reduced semantic processing appeared to have potentiated later memory.

Present Study

In the present study, we tested how prosodic prominence predicted performance in a later memory task: free recall. Naïve dyads first completed a modified version of the map task (Anderson et al., 1991). One participant, the *director*, viewed a route connecting a series of pictures, as in the top panel of Figure 1. The director described the route to a second participant, the *matcher*, who had to draw the route on a copy of the map that contained only the pictures, as in the bottom panel of Figure 1. (1) through (3) present a typical series of instructions given by the director.

(1) Go from the vase to the bone.

(2) Now, go from the star to the kite.

(3) Now, go from the vase to the pants.

Each critical referent was mentioned exactly twice, as in (1) and (3), allowing the second mention to potentially be reduced relative to the first. After the map task, the director completed a surprise written free recall test for the referents described in the map task. Acoustic measurements of the spoken productions in the map task were then sorted based on whether or not the speaker eventually recalled the referent on the free recall test. One advantage of this method is that it permits a within-item comparison between instances when a particular referent

was recalled versus instances when the same referent was not recalled. This within-item comparison is thus orthogonal to item-level variables such as lexical frequency (Bell et al., 2009) or visual complexity (Christodoulou, 2012) that might affect prosodic production or recall.

(INSERT FIGURE 1 ABOUT HERE)

We had two hypotheses about speakers' production of prosodic prominence in this task. First, based on prior findings (e.g., Fowler & Housum, 1987), we hypothesized that the second mention of a referent would be reduced in prominence compared to the first. Second, we hypothesized that, if prosodic reduction is influenced by the ease of retrieving the relevant discourse representations and lexical forms, the *degree* of reduction at the time of the second mention should be predictive of whether the speaker eventually recalls the referent on the final memory test. As described above, theories of human memory postulate that items that require the most effort to retrieve at the time of the second mention—that is, the items that were least prosodically reduced—should benefit the most from the second presentation and thus be most apt to be recalled by the speaker. Such a relation between prosodic prominence and later recall would provide evidence that prosodic prominence is influenced by one of the same factors that influences later recall: ease of retrieval.

The predictions of the processing facilitation model are depicted in Figure 2. First, consider an item that is relatively easy to retrieve at the time of the second mention (top row). A readily available item requires less effortful production and results in reduced prosodic prominence of the second mention. However, such an easy retrieval is dissimilar to the demands of the eventual recall test and thus does not constitute the effortful retrieval practice that potentiates the speaker's later memory (Benjamin & Tullis, 2010; R. Bjork & Allen, 1970;

Cepeda et al., 2006; Finley et al., 2011; Karpicke & Roediger, 2008; Matzen & Federmeier, 2010; Melton, 1967; Tullis, Benjamin, & Ross, in press). This result contrasts with the pattern for an item that is less easy to retrieve at the time of the second mention (bottom row). This low availability results in an effortful, prosodically prominent second mention, and it is exactly such effortful practice that increases the probability of later recall (Benjamin & Tullis, 2010; R. Bjork & Allen, 1970; R. Bjork & E. Bjork, 1992; Tullis, Benjamin, & Ross, in press).

(INSERT FIGURE 2 ABOUT HERE)

In contrast, if acoustic prominence is determined solely by the grammatical constraints of a language (e.g., a word is reduced if it is repeated), as in trigger models, we would expect no relationship between the acoustic prominence of a word and later recall. The target words in this experiment all occur in contexts in which they are all repeated and are all visually copresent for the speaker and addressee. Consequently, a grammatical theory of prominence would predict that these words should be reduced to an equal extent.

Because prosodic prominence may be realized by multiple acoustic cues (Breen et al., 2010; Wagner & Watson, 2010), the acoustic duration, the mean intensity, and the mean fundamental frequency of speakers' productions were all measured; we discuss possible differences between these acoustic measures in the General Discussion.

Method

Participants

Thirty-six dyads of naïve participants participated in the experiment in partial fulfillment of a course requirement. Each of the two participants was recruited individually for the session, and participants were unfamiliar with each other prior to the session. All participants were

undergraduate students at the University of Illinois who reported on a questionnaire that they were monolingual native speakers of American English.

Three dyads were eliminated from analysis because the speaker repeated item names twice in every instruction (e.g., *Start at the skis and go from the skis to the slide*), which meant the second mention always occurred in a highly predictable context. These eliminations resulted in thirty-three dyads in the final analysis.

Materials

We selected 140 black and white line art drawings, 42 of which served as the critical items, from the International Picture Naming Project (Szekely et al., 2004). All selected pictures depicted objects.

Items were divided into seven printed maps, each consisting of 20 pictures arranged in a 4 x 5 array. To encourage participants to use the item names (e.g., *skis*) rather than locations in the grid (*third object in the first row*), items' horizontal and vertical positions were randomly adjusted so that they did not form strict rows or columns. On the *director map*, 12 numbered arrows pointed from one picture to another. Six critical items per map appeared in two of these pairs; another 12 filler items appeared in only one pair. Because words at the end of a prosodic phrase are subject to phrase-final lengthening that may obscure other prosodic effects (for review, Wagner & Watson, 2010), the critical items always appeared as the first picture in each pair. A complete list of pictures and pairings is available in the Appendix.

Two distractor pictures in each map were not touched by any arrows; these items were included so that the items in the last pair were not completely predictable to the matcher. Furthermore, to ensure that the matcher could not predict when the second mention of each item

would occur, the lag between the two instructions referring to each critical item was varied within participants between 0, 1, or 3 intervening instructions[1].

A second version of each map, the *matcher map*, displayed the same pictures in the same arrangement as the director map but without the arrows or numbers.

Procedure

One participant was randomly designated the *director* and the other the *matcher*. The director was given the director maps one at a time and told to describe the sequence of the pictures in order to the matcher, using the example structure *Go from the airplane to the lion*. The matcher was instructed to use these directions to draw in the numbers and arrows on the matcher map.

To ensure that directors used the object names in their directions, directors were instructed not to use relative terms such as *to the left* or *above*. No other restrictions were placed on the form of participants' instructions.

The director and matcher first completed a three-item practice map, followed by the seven critical maps. A Marantz PMD670 Professional digital recorder was used to record the director's speech. To obtain reliable measures of intensity, the director wore a headset microphone that was positioned a constant distance from the director's mouth.

Following the completion of the last map, each participant was given a sheet of paper and told to write down as many of the object names as possible. Participants were not informed about the recall test until after completing the last map, so productions were unlikely to reflect a strategic attempt to prepare for a memory test. Both the director and matcher completed the recall test to make the experimental session an equal length for both participants, but only data

from the director was relevant for evaluating our hypotheses about the relation between prosodic prominence and a speaker's ease of retrieval.

Coding

The first author coded the start and end of each critical word using the Praat speech analysis software (Boersma & Weenink, 2010). To assess the reliability of the coding procedures, six dyads (18% of total trials) were also coded by a second rater. The durations of the critical words as coded by the first rater and as coded by the second rater were highly correlated, $r = .96$, $p < .001$, indicating substantial agreement between the raters in locating the start and end of the words. Where the two raters disagreed, the first rater's coding was used.

Observations were excluded if the director mentioned the items out of order (3%), if the director repeated the item name multiple times because the matcher had difficulty locating the item (3%), if the recording was obscured by external noise such as coughing or tapping on the microphone (1%), or if a pitch track could be not obtained in Praat (2%). Observations were also excluded if the director produced an unexpected name for the critical picture, such as the unexpected name *fish* for a whale picture (8%); in these cases, there were not enough tokens of the unexpected names to compare productions of the same name when it was remembered on the final test versus when it was not. These exclusions left 1148 of the tokens (83%) for analysis.

Using Praat, mean fundamental frequency (F0) was then calculated using autocorrelation, and intensity was obtained from mean energy with mean pressure subtracted. Duration of the spoken words was obtained directly from the coding of when each word began and ended.

Results

Although the director typically recalled both critical and filler referents, only recall of the

critical repeated referents was relevant to our hypothesis that the prosody of repeated mentions would predict their subsequent recall. Figure 3 plots mean intensity (top panel), duration (middle panel), and mean fundamental frequency (F0; bottom panel) of the critical items as a function of mention number (first or second) and whether or not the director eventually recalled the item on the subsequent memory test.

(INSERT FIGURE 3 ABOUT HERE)

We assessed the relation of prosodic prominence to directors' memory for the discourse in two analyses. First, we demonstrated that repeated mentions were reliably reduced in intensity, duration, and F0. Then, we tested facilitation-based accounts of this reduction by assessing whether prosodic reduction predicted the probability that a director could later recall a particular referent.

Reduction

To assess the statistical reliability of prosodic reduction, we modeled the intensity, duration, and F0 of each token using multi-level models, which can capture variability across both subjects and items. Each model included mention number as a mean-centered fixed effect, subject and item as random intercepts, and the maximum random effects structure justified by the data.

The models of prosodic reduction were fit in the R software package with restricted maximum likelihood estimation using the *lmer()* function of the *lme4* package (Bates, Maechler, & Bolker, 2011). Degrees of freedom for the *t* statistics were obtained from the Satterthwaite approximation using the *lmerTest* package.

Intensity model. A likelihood ratio test revealed that the intensity model was improved

by the addition of a random slope of mention number by subjects, $\chi^2_{(2)} = 9.82$, $p < .01$, but not further improved by a random slope by items, $\chi^2_{(2)} = 0.65$, $p = .72$. The model revealed that mean intensity was reliably reduced between mentions; on average, second mentions were 0.48 dB less intense than first mentions, 95% CI: [0.28, 0.70], $t_{(29.66)} = 4.58$, $p < .001$.

Duration model. The duration model was improved by the addition of a random slope of mention number by subjects, $\chi^2_{(2)} = 56.30$, $p < .001$, and further by a random slope by items, $\chi^2_{(2)} = 29.35$, $p < .001$. As with intensity, duration was reliably reduced between mentions: Second mentions were on average 95 ms shorter than first mentions, 95% CI: [75, 115], $t_{(41.92)} = 9.28$, $p < .001$.

F0 model. Several measures of F0 are available, including mean F0, maximum F0, and F0 range. We focused on mean F0 because prior findings (Breen et al., 2010) have indicated this measure had the strongest influence on naïve listeners' perception of prominence, but identical findings were obtained when F0 range or maximum F0 was used instead.

The F0 model was not reliably improved by a random slope by subjects, $\chi^2_{(2)} = 0.89$, $p = .64$, nor by a random slope by items, $\chi^2_{(2)} = 4.57$, $p = .10$. Like the other measures of prominence, mean F0 was reliably reduced between mentions; the mean F0 of second mentions was on average 5.51 Hz lower than that of first mentions, 95% CI: [3.948, 7.08], $t_{(2221)} = 6.89$, $p < .001$.

Summary. Duration, intensity, and F0 all showed reliable reduction from first to second mentions, consistent with past work (e.g., Fowler & Housum, 1987). Moreover, the degree of reduction was generally comparable to prior results in the literature (e.g., 95 ms reduction in duration and 5.5 Hz reduction in F0 in the present study versus 56 ms and 5 Hz in Fowler &

Housum, 1987). Observing this typical reduction in prominence across mentions is crucial because it allows a test of what drives prosodic reduction in this and other studies. In particular, we tested whether reduction relates to the speaker's future recall, as predicted by facilitation-based accounts of prosodic prominence.

Director's Recall

Having established that prosodic prominence was reduced in second mentions, we then assessed whether prominence predicted the director's ability to later recall the referents he or she had mentioned. For this model, the dependent variable was binary (each referent was either recalled by the director or not); thus, we modeled the log odds that the director would recall a particular referent as a function of the acoustic correlates of prominence. This logistic model was fit with Laplace estimation using *lme4* (Bates et al., 2011); for such models, Wald *z* statistics are directly available. We centered each of the predictor variables around its mean to obtain a more meaningful intercept term (representing recall performance at an average level of prosodic prominence). Because there were no effects of lag and because lag was a counterbalancing variable unrelated to the questions of interest, the results will be presented with the data collapsed across lag conditions. Including lag as a factor did not change any of the results.

Of primary theoretical interest was whether the director's recall of the critical referents was predicted by their reduction[2] (or lack thereof) across mentions in mean intensity (in dB), in duration (in ms), or in mean F0 (in Hz), as predicted by facilitation-based theories. However, it was important to establish that any role of prosodic *reduction* in predicting recall was not due to a confound with the prominence of the *initial* mention. To rule out such a possibility, we first fit a model that included only the mean intensity, duration, and mean F0 of the initial mention. The

results of this reduced model are presented in Table 1 and indicate no relation between initial prominence and subsequent recall. With this control in place, we then added the critical measures of prosodic reduction[3] from the first mention to the second mention; as will be seen, prosodic reduction *did* predict eventual recall.

(INSERT TABLE 1 ABOUT HERE)

Note that the fit of neither the reduced model nor the final model was improved by random slopes, as indicated in likelihood-ratio tests (all *p*s > .10); we thus report the model with only random intercepts because this was the maximum random effects structure justified by the data. Including any of the random slopes did not change any of the conclusions presented below. Parameter estimates and their correlations for the final model are displayed in Tables 2 and 3.

(INSERT TABLE 2 ABOUT HERE)

(INSERT TABLE 3 ABOUT HERE)

Reduction in intensity was reliably related to the director's ability to later recall the referents. Figure 4 plots the probability that the director eventually recalled a referent as a function of reduction in mean intensity of the referring expressions. For each 1 dB reduction in mean intensity, the odds of later recall in the model were 0.93 times lower, 95% CI [0.86, 0.99], $z = -2.15$, $p < .05$. This pattern is consistent with the claim that repeated study is less beneficial for items that are already easy to retrieve, and thus prosodically more reduced, at the time of the second presentation (Benjamin & Tullis, 2010; R. Bjork & E. Bjork, 1992; R. Bjork & Allen, 1970).

(INSERT FIGURE 4 ABOUT HERE)

Reduction in duration was not related to the director's subsequent recall, $p = .98$, nor was

reduction in F0, $p = .62$. Recall probability is plotted as a function of reduction in duration in Figure 5 and as a function of reduction in F0 in Figure 6; neither plot reveals a systematic relationship. We return to the difference between intensity and the other acoustic measures in the General Discussion.

(INSERT FIGURE 5 ABOUT HERE)

(INSERT FIGURE 6 ABOUT HERE)

Discussion

Words in running speech vary in their prosodic prominence, with repeated or predictable words typically reduced in their prominence (Aylett & Turk, 2004; Fowler & Housum, 1987; Jurafsky et al., 2001). One hypothesized source of this reduction is the ease with which the language production system can retrieve the relevant discourse representations or lexical forms. Referents and lexical items that are new or unpredictable require effortful, prosodically prominent productions. However, repetition and predictability facilitate the retrieval of particular discourse representations or lexical forms, allowing them to be produced quickly and with reduced prominence.

In the present study, we tested such facilitation-based accounts of prosodic prominence by examining the relation of prosodic prominence to a speaker's subsequent recall of the referents in the discourse. Theories of human memory indicate that a second study opportunity better potentiates later memory when items are less available at the time of the second study and thus require effortful retrieval (Benjamin & Tullis, 2010; R. Bjork & Allen, 1970; R. Bjork & E. Bjork, 1992). If prosodic reduction stems at least in part from facilitated retrieval, items *least* reduced in prominence (and, by implication, *most* effortful to retrieve) at the time of the second

presentation should later be best remembered by the speaker. This prediction was supported for the acoustic measure of intensity.

Moreover, Figure 4 reveals that this relationship between intensity and recall was continuous: Additional changes in reduction led to additional changes in probability of recall by the speaker. This pattern is unlikely to be explained by speakers' deliberate choice of one linguistic form or another because speakers would have to be choosing among numerous, incrementally different forms of each word. Nor can the pattern be explained exclusively as noise in realizing a simpler binary distinction between prominent and non-prominent forms. Although it is plausible that there would be some noise in realizing such a distinction, the variability in the present study was not mere noise; it actually *predicted* subsequent recall, suggesting it was related to speaker's cognitive processes. Rather, the data are consistent with the predictions of an account in which continuous variation in ease of retrieval within the language production system results in continuous variation in prosodic prominence.

These results do not imply that facilitation in the language production system is the *only* source of prosodic prominence. Prosodic prominence might actually reflect facilitation summed with deliberate choices on the part of the speaker between different forms, or prosodic prominence in other situations might purely reflect categorical choices. Speakers might choose, for instance, between accented forms and deaccented forms to indicate whether information is new to the discourse versus already given in the discourse (Dahan, Tanenhaus, & Chambers, 2002; Pierrehumbert & Hirschberg, 1990; Schwarzchild, 1999; Selkirk, 1995, 2002; Steedman, 2002), or between different accented forms to denote new versus contrastive information (Fraundorf, Watson, & Benjamin, 2010, 2012; Pierrehumbert & Hirschberg, 1990; Selkirk, 1995,

2002; Steedman, 2002). Facilitation-based accounts also do not preclude prosodic prominence being a useful cue for listeners, as has been observed in language comprehension experiments (e.g., Dahan et al., 2002). Listeners could use the correlation between facilitated production and prominence to infer whether a word is likely to refer to a more or less available item even if speakers do not deliberately choose to produce prominent forms.

The facilitation-based account of the present data proposes that a common variable—ease of retrieval at the time of the second mention—underlies variation in both prominence and recall performance. A somewhat different explanation is that the act of speaking a word with greater intensity is the actual causal mechanism that facilitated later memory for that word. Although it is not completely possible to rule out such an account, the data provide evidence against it. Speaking a word with greater intensity did *not* always lead to a higher probability of recall; as previously noted, the intensity of the first mention alone showed no relation to later memory. Had the relation of prosodic prominence to recall been driven by articulation itself, there is no reason why a second mention should have a special role in predicting later memory. Rather, the key predictor was how intensity changed when the item was presented a second time and could potentially be retrieved from memory.

One other alternate account is that the observed relationship between prosodic prominence and later recall simply reflects item-specific characteristics, such as lexical frequency (Bell et al., 2009) or visual complexity (Christodoulou, 2012), that happen to affect both prosodic production and explicit recall. This possibility can be ruled out as an account of the present data because the crucial comparisons in this study were within-item; tokens of the same word being produced in response to the same picture (e.g., *vase*) could be compared when

the item was later remembered versus later forgotten. This comparison is independent of item-level properties. Moreover, the statistical models provided no evidence that the effects of interest varied across items. In principle, variability in an effect across items can be accounted for in a multi-level model by a random slope of that effect by items. However, random slopes did not reliably contribute to the model of intensity reduction across mentions nor to the model of later recall; moreover, including such slopes did not change the effects of interest. Thus, there was no evidence that the critical effects were driven by the properties of particular items.

Given that the critical comparisons were within-items, it might be asked what led a particular referent to be remembered by some participants and forgotten by others—for instance, why one participant remembered *vase* but another participant forgot it. Some theories of memory (e.g., Wixted, 2007) have proposed that there should indeed be significant variability in how much initial exposures potentiate later memory. For example, individuals may relate some referents but not others to their prior knowledge, which typically enhances memory (Craik & Lockhart, 1972), or they may devote more time and attention to some referents (Tullis & Benjamin, 2010). Referents may also differ across subjects in their idiosyncratic memorability; for example, high-frequency items are easier to recall, but individuals' differing experiences may make some referents (or words) more or less frequent and thus more or less memorable (Worden & Sherman-Brown, 1983). The present work was not designed to enumerate all such variables influencing initial memory; rather, our goal was to leverage this variability to test theories of prosodic prominence.

The present study concerned the mechanisms underlying variation in prominence. A related, but separate question in the literature is whether speakers produce this variation in

prosodic prominence "for the listener"—that is, with consideration of the addressee's needs rather than as a consequence of the speaker's own processing needs. Some evidence suggests that the effect of interest was not produced for the listener's benefit: The relatively small difference between recalled and unrecalled words (M = 0.4 dB) is one that listeners are unlikely to perceive or use to constrain their interpretation, and there was indeed no evidence that the speaker's prosodic prominence influenced the matcher's recall[4]. However, the present study was not specifically designed to address questions of audience design. When a referent was repeated, it had already been seen by both the director and matcher; thus, it is unclear whether the addressee's experience with the referent had an effect on prosodic prominence separate from the director's experience. Moreover, even if the directors tailored their prosodic prominence to the matcher, it is not clear that this is a cue that would necessarily influence the matcher's subsequent recall. Thus, it remains an open question as to the degree to which speakers produce variation in prosodic prominence for the benefit of the listener (but see Arnold, Kahn, & Pancani, 2012, for some evidence that prosodic prominence can be influenced by audience design).

Multiple Sources of Prosodic Prominence

In the present study, variation in prominence predicted later explicit memory, but this relation was carried by acoustic intensity; no such relation was observed for variation in duration or fundamental frequency (F0). Some possible accounts of this difference are that reduction in duration and in F0 is less reliable or that duration and F0 are simply insensitive to the processes that underlie prosodic reduction. However, second mentions were indeed reliably reduced in both duration and F0 in this study, as they have been in numerous other studies (e.g., Bard et al., 2000; Fowler & Housum, 1987; Lam & Watson, 2010, 2014; Vajrabhaya & Kapatsinski, 2011);

rather, the degree of reduction was unrelated to later recall.

(INSERT TABLE 4 ABOUT HERE)

The acoustic correlates of prominence may have had differing relations to later explicit recall because they reflect different processes in language production. Within the critical words (which were always in relatively prominent positions), the three acoustic correlates of prominence of interest were only weakly or even negatively correlated, as shown in Table 4. The differences among acoustic measures is consistent with some theoretical accounts; in particular, Watson (2010) has proposed a *multiple source account* of prosodic prominence in which multiple processes contribute to prosodic prominence and may thus differentially affect particular acoustic measures. Indeed, several studies have found that duration and intensity show different influences and do so in ways consistent with the present data. Several experiments (Lam & Watson, 2010, 2014; Ouyang & Kaiser, in press) have found reduction in duration stems mostly from lexical repetition; reduction of duration effects has thus been attributed (Bard et al., 2009; Lam & Watson, 2014) to implicit activation of lexical forms. However, such implicit memory effects rarely transfer across modalities or to free recall (Roediger, 1990), a task that primarily probes participants' knowledge of what referents were discussed in the discourse rather than their ability to quickly produce lexical forms. Thus, form-level processes reducing *duration* would have been unlikely to influence performance on the free recall task in the present study. By contrast, discourse-level manipulations such as referential predictability (Lam & Watson, 2010; 2014), contrastiveness (Ouyang & Kaiser, in press), and importance in a conversational game (Watson, Arnold, & Tanenhaus, 2008) affect primarily *intensity*; such discourse-level representations would also have been essential for recalling the referents of the discourse in the

free recall task. The divergence between duration and intensity in their relation to free recall is thus consistent with prior evidence that duration is most strongly affected by form-level processing but intensity by discourse-level processing. It is less clear why variation in fundamental frequency (F0) did not predict later recall, but one possibility raised by Arnold and Watson (in press) is that F0 is most reflective of the speaker's choice of particular linguistic forms, such as a choice among intonational contours that may carry different meanings (Gussenhoven, 1983; Pierrehumbert & Hirschberg, 1990; Selkirk, 2002; Steedman, 2000), rather than the facilitation of language production.

Assessing Language Production Through Future Memory

The present results provide data that converge with results from scene description and event description tasks (e.g., Kahn & Arnold, 2012, in press; Lam & Watson, 2010, 2014; Ouyang & Kaiser, in press; Vajrabhaya & Kapatsinski, 2011) in suggesting that prosodic prominence is influenced by the ease of retrieving the representations required for language production—and that different acoustic correlates of prominence may be influenced by different variables. One potential concern with conclusions from those tasks is that some (though not all) of the tasks, especially those purporting to demonstrate multiple sources of prosodic prominence, involved experimental manipulations that disrupted the natural correlations between variables. For example, Lam and Watson (2010) separately manipulated lexical repetition and referential predictability. But in natural conversation, repetition and reference are positively correlated: Referents that have already been mentioned are more likely to be mentioned again (Arnold, 1998). Event description tasks that disrupted this prior distribution might have led to within-experiment learning that shifted participants' production preferences. (For one such example,

see Lam & Watson, 2014.)

More naturalistic evidence in favor of facilitation-based account of reduction has come from corpora of spoken speech (Bell et al., 2009; Gahl, Yao, & Johnson, 2012). However, as with all studies in which materials are not under laboratory control, it is possible that effects in a corpus analysis attributed to the availability of lexical or discourse representations instead reflect some other uncontrolled and unmeasured item-level property.

An advantage of the present paradigm is that it allows controlled, within-item comparisons but does not require artificial orthogonal manipulations. Comparisons were based on the natural tendency of participants to remember items in some instances and forget them in others. These comparisons revealed that repeated mentions of the same when it was eventually remembered versus when it was eventually forgotten differed in their prosodic prominence and did so in ways consistent with the claim that prosodic prominence partially reflects the difficulty of retrieval. These converging results both support facilitation-based accounts of prosodic prominence and validate the scene and event description tasks as generally assessing naturalistic prosody.

Conclusion

Prosodic prominence can predict later explicit recall of a discourse: Repeated referring expressions that were least reduced in acoustic intensity were most apt to later be recalled by the speakers. Given that repetition better potentiates later memory for items that require more difficult retrieval at the time of the repetition (Benjamin & Tullis, 2010; R. Bjork & Allen, 1970; R. Bjork & E. Bjork, 1992; Tullis, Benjamin, & Ross, in press), this pattern supports facilitation-based accounts of prosodic reduction. Items that were most difficult to retrieve when re-

presented (as inferred through later recall) were the least prosodically reduced, but the items that were apparently easy to retrieve at the time of the re-presentation were more prosodically reduced. One source of variation in prosodic prominence may thus be continuous variation in the ease of retrieving representations and producing language.

Acknowledgements

We thank members of the Communication and Language, Language Production, and Human Memory and Cognition Labs at the University of Illinois for their comments and suggestions, and Shelby Luzzi, Amie Roten, Erika Schmit, and Samuel Weeks for data collection and transcription.

References

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., A., Doherty, G., Garrod, S., . . . Weinert, R. (1991). The HCRC Map Task corpus. *Language and Speech*, *34*, 351-366. doi:10.1177/002383099103400404

Arnold, J. E. (1998). *Reference form and discourse patterns*. Unpublished doctoral dissertation, Stanford University.

Arnold, J. E., Kahn, J. M., & Pancani, J. C. (2012). Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin & Review*, *19*, 505-512. doi: 10.3758/s13423-012-0233-y

Arnold, J. E., & Watson, D. G. (in press). Synthesizing meaning and processing approaches to prosody. *Language, Cognition and Neuroscience.* doi: 10.1080/01690965.2013.840733

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*, 31-56. doi:10.1177/00238309040470010201

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, *42*, 1-22. doi:10.1006/jmla.1999.2667

Bates, D., Maechler, M., & Bolker. B. (2011). lme4: Linear mixed-effects models using s4 classes [Computer software]. Retrieved from http://CRAN.R-project.org/package=lme4 (R package version 0.999375-42)

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on

durations of content and function words in conversational English. *Journal of Memory and Language*, *60*, 92-111. doi:10.1061/j.jml.2008.06.003

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology*, *61*, 228-247. doi:10.1016/j.cogpsych.2010.05.004

Bjork, R. A., & Allen, T. W. (1970). The spacing effect: Consolidation or differential encoding? *Journal of Verbal Learning and Verbal Behavior*, *9*, 567-572. doi: 10.1016/S0022-5371(70)80103-7

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *Learning processes to cognitive processes: Essays in Honors of William Estes* (Vol. 2, pp. 35-67). Hillsdale, NJ: Erlbaum.

Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer [Computer program]. Version 5.1.30, retrieved 25 March 2010 from http://www.praat.org

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*, 1044-1098. doi: 10.1080/01690965.2010.504378

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, *34*, 268-276. doi: 10.3758/BF03193405

Cepeda, N. J., Pashler, H., Vul, E. Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and qualitative synthesis. *Psychological Bulletin, 132*, 354-380. doi:10.1037/0033-2909.132.3.354

Christodoulou, A. (2012). *Variation in word duration and planning.* (Unpublished doctoral dissertation.) The University of North Carolina at Chapel Hill, Chapel Hill, NC.

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonetics*, *1*, 425-452. doi: 10.1515/LABPHON.2010.022

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning Verbal Behavior*, *11*, 671-684. doi: 10.1016/S0022-5371(72)80001-X

Crowder, R. G. (1976). Principles of learning and memory. Oxford, England: Lawrence Erlbaum.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*, 292-314. doi: 10.1016/S0749-596X(02)00001-3

Finley, J. R., Benjamin, A. S., Hays, M. J., Bjork, R. A., & Kornell, N. (2011). Benefits of accumulating versus diminishing cues in recall. *Journal of Memory and Language*, *64*, 289-298. doi:10.1016/j.jml.2011.01.006

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, *26*, 489-504. doi:10.1016/0749-596X(87)90136-7

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

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66, 789-806. doi: 10.1016/j.jml.2011.11.006

Griffin, Z., & Ferreira, V. S. (2006). Properties of spoken language production. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (2nd ed.; pp. 21-59). London: Elsevier.

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

Jurafsky, D., Bell, A., Gregory, M. & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee and P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229-254).

Kahn, J. M., & Arnold, J. E. (2012). A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language*, *67*, 311-325. doi: 10.1016/j.jml.2012.07.002

Kahn, J. M., & Arnold, J. E. (in press). Articulatory and lexical repetition effects on durational reduction: speaker experience vs. common ground. *Language, Cognition and Neuroscience.* doi: 10.1080/01690965.2013.848989

Karpicke, G., & Roediger, H. L. III (2008). The critical importance of retrieval for learning. *Science*, *319*, 966. doi: 10.1126/science.1152408

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, *118*, 1038-1054. doi: 10.1121/1.1923349

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621-647. doi: 10.1146/annurev.psych.093008.131123

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package) [Computer software]. Retrieved from http://CRAN.R-project.org/package=lmerTest (R package version 2.0-6).

Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced less prominent. *Memory & Cognition*, *38*, 1137-1146. doi:10.3758/MC.38.8.1137

Lam, T. Q., & Watson, D. G. (2014). Repetition reduction: Lexical repetition in the absence of referent repetition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 829-843. doi: 10.1037/a0035780

Matzen, L. E., & Federmeier, K. D. (2010, June). Repetitions and reminding: A novel analysis of the Dm effect. Paper presented at the Interbrain ICA Conference, Juvaskyla, Finland, June 12-17.

Melton, A. W. (1967). Repetition and retrieval from memory. *Science*, *158*, 532.

Ouyang, I. C., & Kaiser, E. (in press). Prosody and information structure in a tone language: an investigation of Mandarin Chinese. *Language, Cognition and Neuroscience*. doi: 10.1080/01690965.2013.805795

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

Roediger, H. L. III. (1990). Implicit memory: Retention without remembering. *American Psychologist*, *45*, 1043-1056. doi:10.1037/0003-066X.45.9.1043

Schwarzschild, R. (1999). Givenness, AvoidF, and other constraints on the placement of accent. *Natural Language Semantics*, *7*, 141-177. doi: 10.1023/A:1008370902407

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

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

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

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., . . . Bates, E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language*, *51*, 247-250. doi:10.1016/j.jml.2004.03.002

Tullis, J., Benjamin, A. S., & Ross, B. H. (in press). The reminding effect: Presentation of associates enhances memory for related words in a list. *Journal of Experimental Psychology: General*. doi: 10.1037/a0036036

Vajrabhaya, P., & Kapatsinski, V. (2011). There is more to the story: First-mention lengthening in Thai interactive discourse. *Proceedings of the International Congress of Phonetic Sciences*, *17*, 2050-2053.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody. *Language and Cognitive Processes*, *25*, 905-945. doi:10.1080/01690961003589492

Watson, D. G. (2010). The many roads to prominence: Understanding emphasis in conversation. In B. Ross (Ed.), *The Psychology of Learning and Motivation, Vol. 52* (pp. 163-183). Elsevier. doi: 10.1016/S0079-7421(10)52004-8

Watson, D. G., Arnold, J. E., & Tanenhaus, M. K. (2008). Tic tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition*, *106*, 1548-1557. doi:10.1016/j.cognition.2007.06.009

Wheeler, M. A., & Roediger, H. L. III (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, *3*, 240-245. doi: 10.1111/j.1467-9280.1992.tb00036.x

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition. *Psychological Review*, *114*, 152-176. doi: 0.1037/0033-295X.114.1.152

Worden, P. E., & Sherman-Brown, S. (1983). A word-frequency cohort effect in young versus eldery adults' memory for words. *Developmental Psychology*, *19*, 521-530. doi: 10.1037/0012-1649.19.4.521

Footnotes

[1] Note that because there were two pictures per instruction, lags of 0, 1, or 3 intervening *instructions* translate to 1, 3, or 7 intervening *picture names*.

[2] The same pattern of results obtained when the measure was percentage reduction rather than the simple difference in prominence between the first and second mentions.

[3] The same pattern of results was also observed if only the measures of prosodic reduction were included in the model (excluding initial prominence) or if the prominence of the second mention was used in place of the degree of prosodic reduction.

[4]Although our theoretical motivation was to test facilitation-based accounts of prosodic prominence by examining whether a speaker's prosodic prominence predicted his or her ability to remember an item in the future, and thus our hypotheses concerned the director's recall, we also had the matcher complete the free recall task to equate the session lengths. A multi-level model of the matcher's recall that included the same variables used to analyze the director's recall did not find that any measures of the director's prosodic prominence predicted the matcher's recall, all ps > .15

Table 1

Fixed Effect Estimates (Top), Fixed Effect Correlations (Middle), and Variance Estimates (Bottom) for Multi-Level Logit Model of Director's Recall, Including Initial Prominence Only (N = 1148, log-likelihood: -642.0).

Fixed effect	Coefficient	*SE*	Wald *z*	*p*
Intercept	-1.030	0.182	-5.59	<.001
Mean intensity of first mention (dB)	0.002	0.013	0.39	.70
Duration of first mention (100 ms)	-0.046	0.060	-0.76	.45
Mean F0 of first mention (Hz)	0.003	0.002	1.17	.24
Random effect	s^2			
Subject	0.20			
Item	0.91			

Table 2

Fixed Effect Estimates (Top),and Variance Estimates (Bottom) for Final Multi-Level Logit Model of Director's Recall (N = 1148, log-likelihood: -639.5).

Fixed effect	Coefficient	*SE*	Wald *z*	*p*
Intercept	-1.030	0.182	-5.65	<.001
Mean intensity of first mention (dB)	0.005	0.013	0.39	.70
Reduction between mentions in intensity (dB)	-0.077	0.036	2.15	.03
Duration of first mention (100 ms)	-0.057	0.081	-0.70	.48
Reduction between mentions in duration (100 ms)	0.001	0.072	0.02	.98
Mean F0 of first mention (Hz)	0.004	0.003	1.37	.17
Reduction between mentions in F0 (100 ms)	-0.002	0.003	-0.49	.62
Random effect	s^2			
Subject	0.19			
Item	0.92			

Table 3

Fixed Effect Correlations for Final Multi-Level Logit Model of Director's Recall.

	Correlations					
Measure	1	2	3	4	5	6
1. Intercept	—					
2. Initial intensity	-.002	—				
3. Reduction in intensity	.021	-.090	—			
4. Initial duration	.004	.130	< .001	—		
5. Reduction in duration	-.002	-.013	.103	-.666	—	
6. Initial F0	-.006	-.029	.022	-.003	-.017	—
7. Reduction in F0	.001	-.025	-.137	-.003	-.065	-.416

Table 4

Average correlations among measures of prosodic prominence.

	Correlations				
Measure	1	2	3	4	5
1. Intensity (first mention)	—				
2. Intensity (second mention)	.65***	—			
3. Duration (first mention)	-.17***	-.10*	—		
4. Duration (second mention)	-.07*	.12**	.36***	—	
5. Fundamental frequency (F0 (first mention)	.01	-.08	.15***	.06	—
6. Fundamental frequency (F0) (second mention)	-.06	-.01	.04	.14**	.31***

Note. To control for baseline differences in overall prominence across subjects, correlations were computed within subjects and then averaged across subjects ($N = 33$).

*$p < .05$. **$p < .01$. ***$p < .001$.

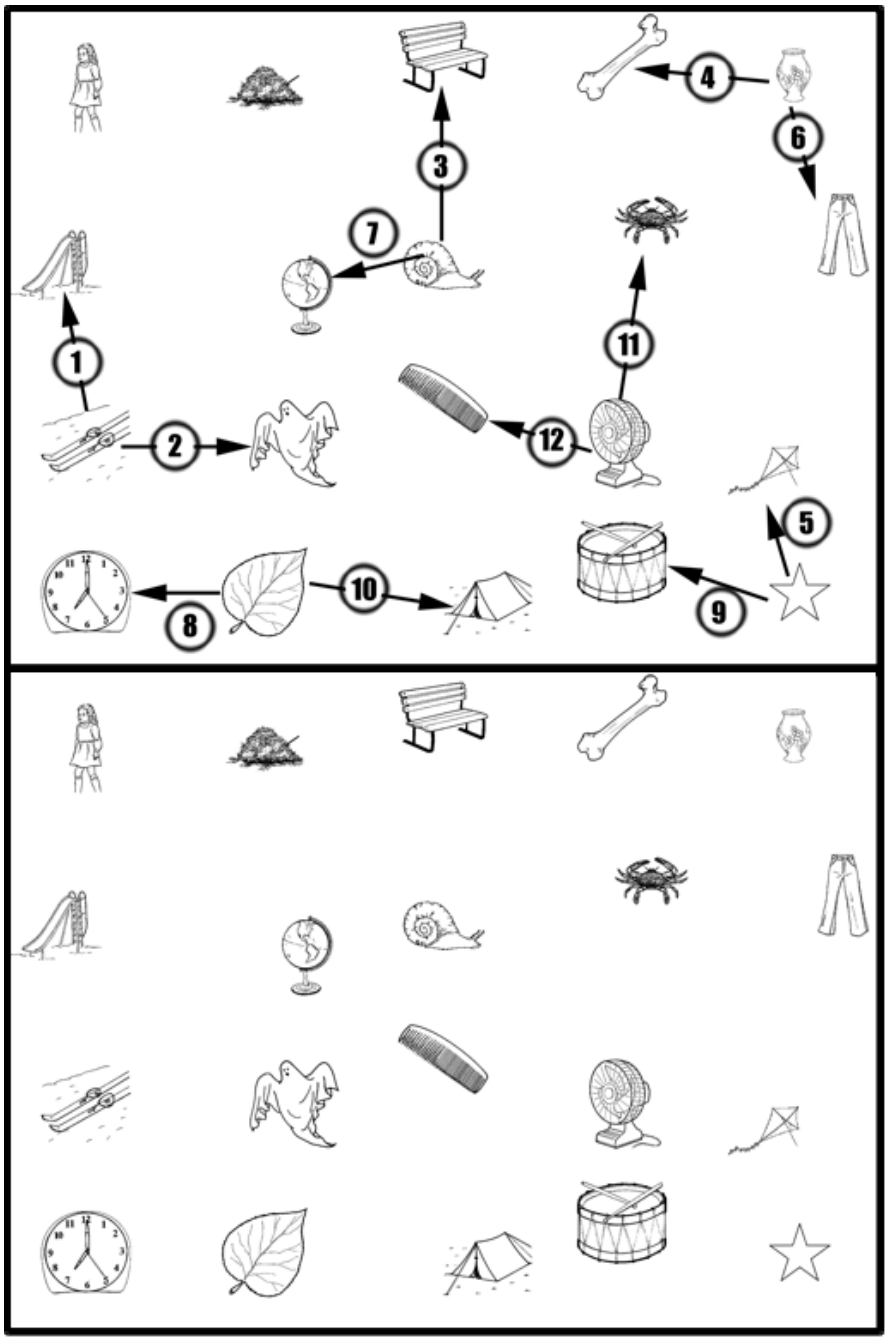

Figure 1. Example map stimuli for director (top panel) and matcher (bottom panel).

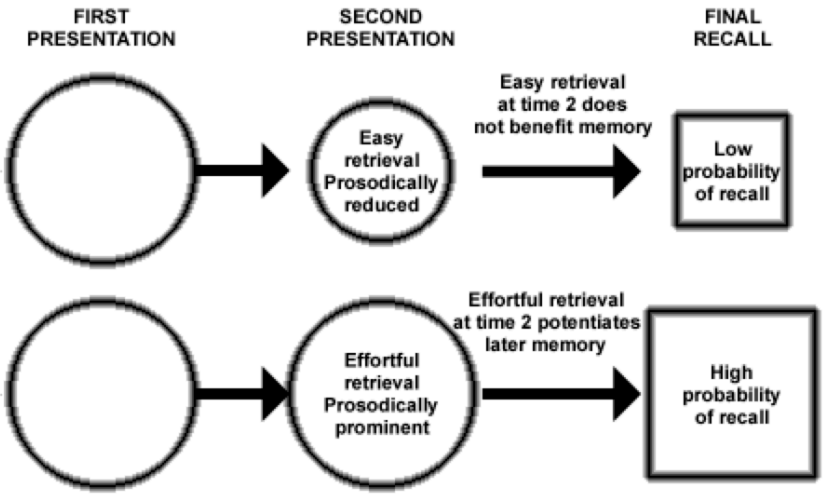

Figure 2. Depiction of predictions from the facilitation-based account for items easy to retrieve (top row) and difficult to retrieve (bottom row) at the time of the second presentation.

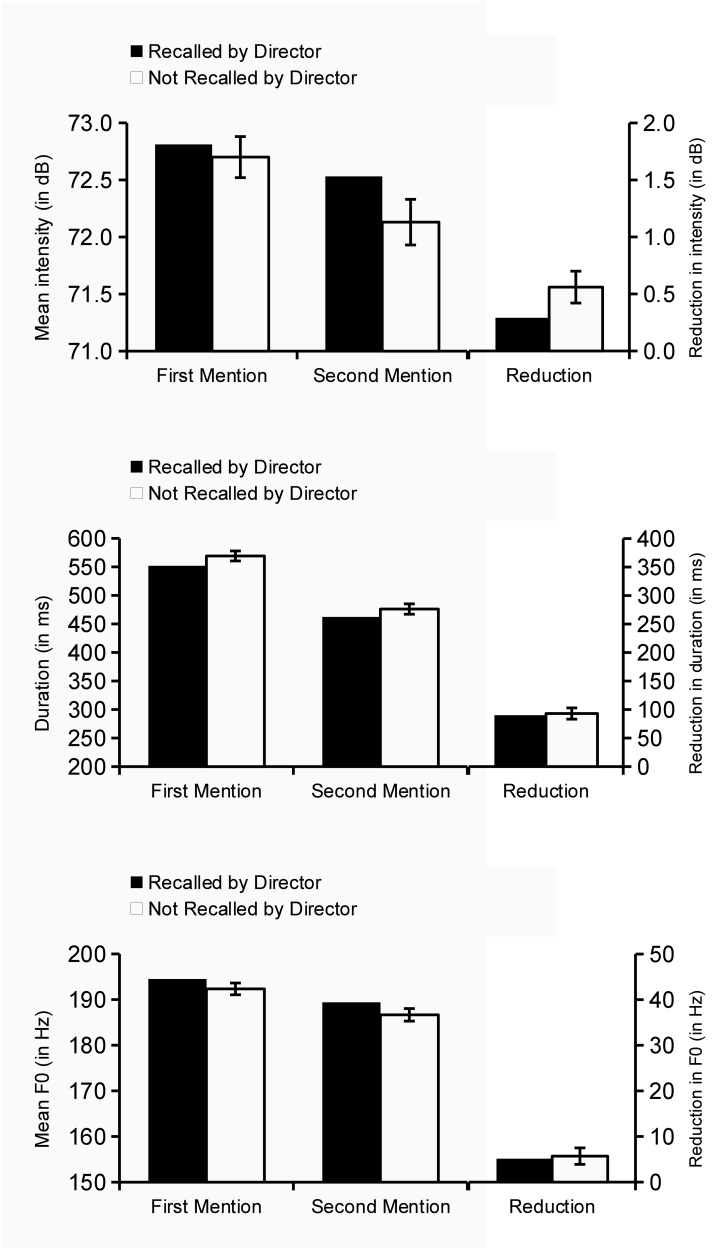

Figure 3. Mean intensity (top panel), duration (middle panel), and mean fundamental frequency (F0; bottom panel) of critical words, as a function of eventual recall by the director, for first mentions, second mentions, and reduction (first mention minus second mention). Error bars indicate the standard error across subjects of the difference between referents that were recalled and referents that were not recalled.

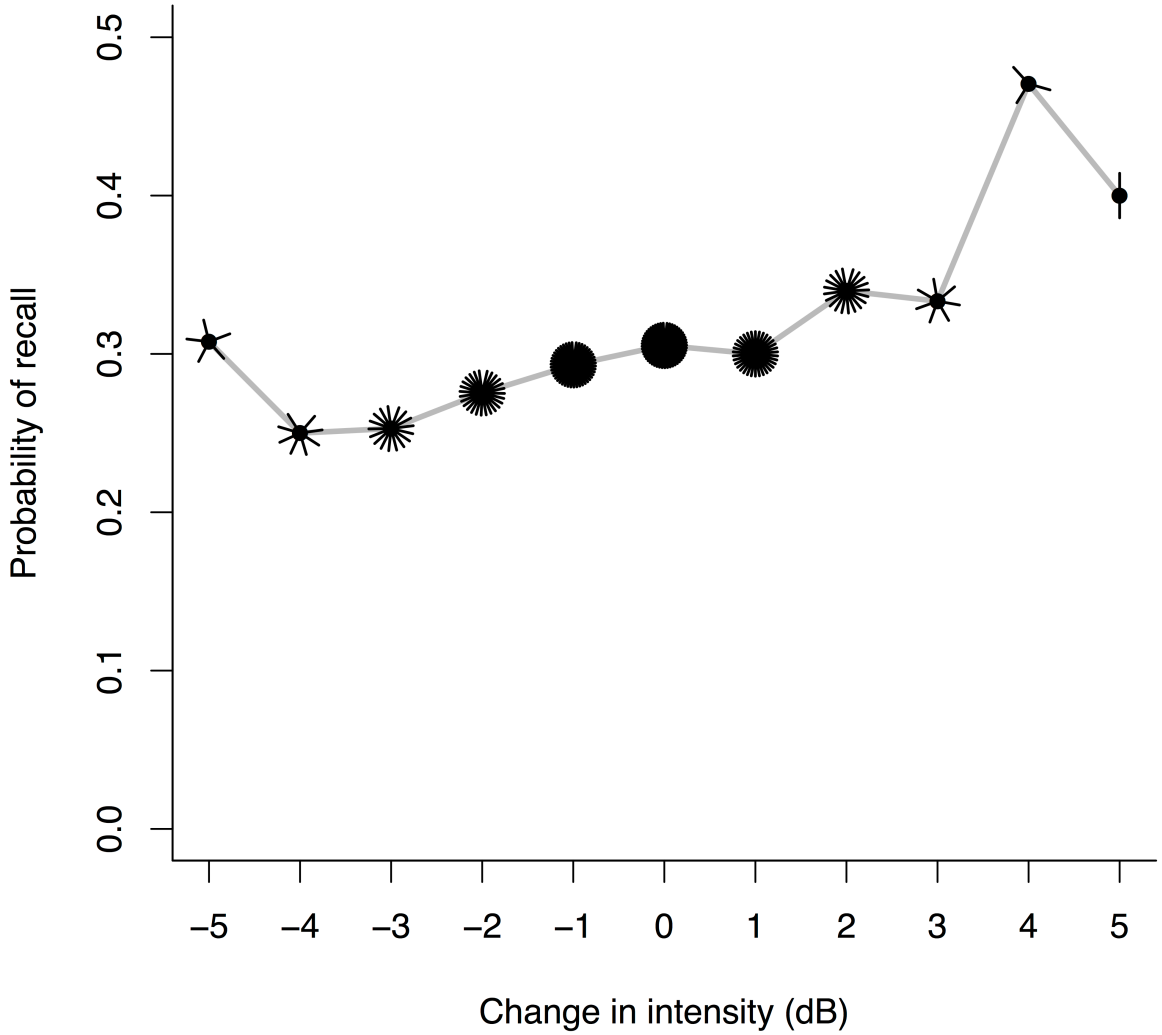

Figure 4. Recall probability as a function of change between mentions in intensity, in 1 dB bins. Negative changes in intensity indicate reduction. The number of rays at each point represents the number of observations in that bin; each ray indicates 5 observations. Bins with fewer than 10 observations are not plotted.

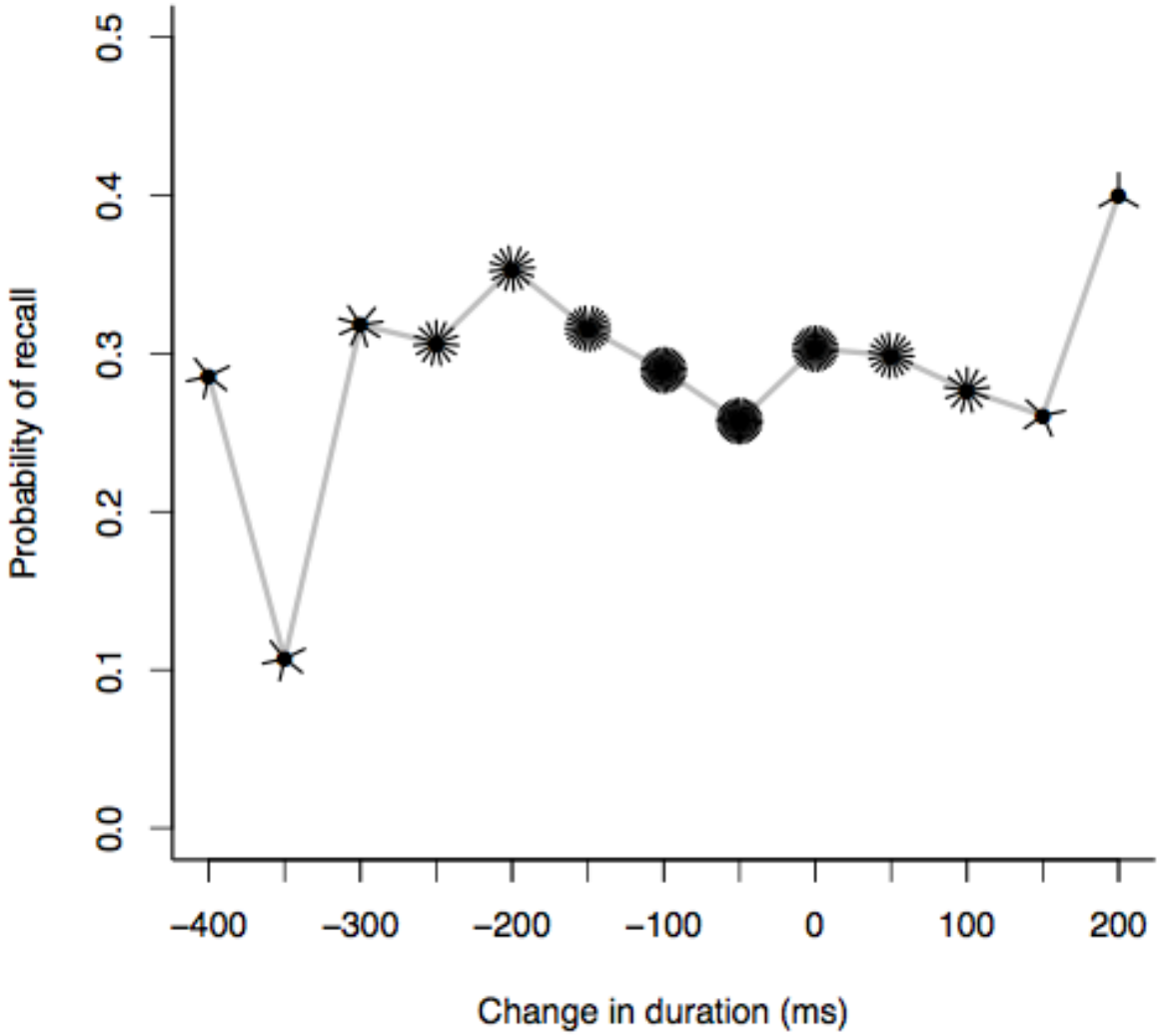

Figure 5. Recall probability as a function of change between mentions in duration, in 50 ms bins. Negative changes in duration indicate reduction. The number of rays at each point represents the number of observations in that bin; each ray indicates 5 observations. Bins with fewer than 10 observations are not plotted.

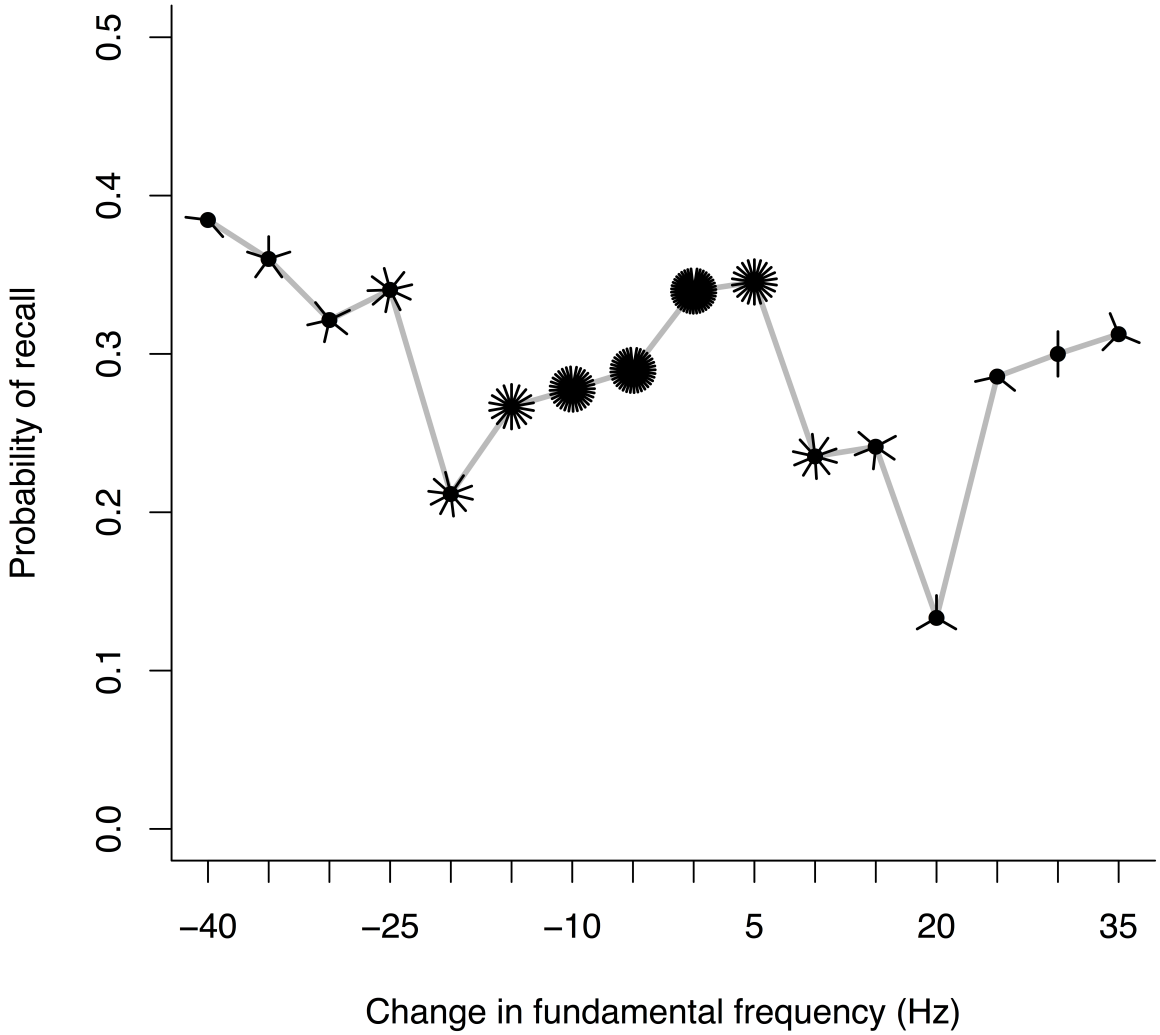

Figure 6. Recall probability as a function of change between mentions in mean fundamental frequency (F0), in 5 Hz bins. Negative changes in F0 indicate reduction. The number of rays at each point represents the number of observations in that bin; each ray indicates 5 observations. Bins with fewer than 10 observations are not plotted.

Appendix

Table A1

List of Target, Filler, and Unmentioned Distractor Pictures.

Mentioned					
	Filler			Filler	
Target	First mention	Second mention	Target	First mention	Second mention
Airplane[a]	Lion[a]	—[b]	Snail	Bell	Harp
Apple[a]	Scissors[a]	—[b]	Saw	Broom	Tent
Lightbulb[a]	Triangle[a]	—[b]	Scarf	Desk	Bench
Skis	Slide	Ghost	Seal	Hook	Church
Axe	Snake	Fork	Chair	Gun	Crown
Stool	Bone	Book	Whale	Frog	Truck
Star	Knight	Drum	Lips	Mop	Shell
Leaf	Ball	Grapes	Ear	Clock	Clown
Well	Pig	Ant	Fox	Sock	Boat
Fence	Bowl	Bus	Witch	Bricks	Tree
Thumb	Branch	Door	Fan	Crab	Globe
Fish	Bag	Worm	Scale	Tank	Key
Watch	Cake	Bat	Swan	Net	Flag
Vase	Kite	Foot	Skull	Leg	Dog
Hose	Sheep	Corn	Mouse	Cow	Glove
Nose	Pen	Car	Eye	Hat	Pot
Chain	Screw	Heart	Swing	Shirt	Pants
Owl	Iron	Lamp	Wrench	Ring	Bomb
Sun	Can	Arm	Wheel	Tie	Jar
Cheese	Wolf	Fly	Moon	Map	Bear
Shoe	Lock	Cat	House	Mask	Bread
Horse	Glass	Bed	Moose	Hand	Bird
Spoon	Knife	Comb			
Unmentioned					
Ladder[a]	Cup	Hay	Stairs	Salt	Couch
Rain	Man	Wheat	Egg	Girl	King
Hinge	Rope	Stove			

[a]Used on the practice map. [b]Practice pictures were mentioned only once.