Supplement Article

Masked Repetition Priming in Treatment of Anomia: A Phase 2 Study

JoAnn P. Silkes^a

Purpose: Previous research has demonstrated that exposure to masked primes may improve naming accuracy for individuals with anomia. This study investigates the effect of repeated exposures to masked identity primes paired with pictures over multiple trials, sessions, and days on the ability of people with anomia to name those pictures.

Method: Four participants with anomia completed this single-subject, multiple-baseline design study. Twelve treatment sessions were conducted for each of 2 semantic categories. Comparisons of performance on naming probes were made between items that were primed, unprimed but seen the same number of times, and unprimed and seen only during naming probes.

Results: All participants showed some gains in naming trained items although to varying degrees, and trained (primed) items generally showed greater improvement than untrained items seen the same number of times. Cross-category generalization was observed for some participants, but little to no within-category generalization occurred. Minimal changes occurred on measures of general language ability.

Conclusions: These data provide continued evidence that masked repetition priming can have a positive effect on naming for people with anomia. Factors that may influence participant response and additional questions that must be settled for this line of research to continue are discussed.

The language behaviors that define aphasia are overt, but the underlying processes that are impaired in aphasia are covert, or implicit—that is, they are largely outside of conscious awareness or control (Tyler, 1992). In terms of an automatic spreading activation approach to understanding lexical processing in aphasia (Dell, 1986; Foygel & Dell, 2000), word retrieval failure is thought to be due to poor activation transmission across levels of lexical representation or poor maintenance of activated linguistic nodes so that they are not active enough at the time of selection (Schwartz, Saffran, Bloch, & Dell, 1994). In terms of a parallel distributed processing model of language (Nadeau, 2001), word retrieval failure is thought to be due to impaired coactivation of the various relevant elements in the language networks. In either case, the impaired ability to retrieve words is related to a failure of automatic, implicit activation processes.

All language tasks—and, therefore, all language treatment tasks—recruit implicit language processes and representations. With this in mind, aphasia treatments and, to be specific, treatments for anomia, generally involve

heavily explicit tasks that require conscious consideration and manipulation of linguistic material, such as semantic and phonological features (e.g., Boyle, 2004; Boyle & Coelho, 1995; Kendall et al., 2008; Linebaugh, Shisler, & Lehner, 2005). These impairment-based treatment approaches are designed to ultimately change implicit processes, and it is clear that these primarily explicit approaches can make positive changes in lexical retrieval for people with aphasia. There is evidence, however, that individuals with aphasia often demonstrate implicit language competence that cannot be expressed in explicit tasks (Andreewsky & Seron, 1975; Mimura, Goodglass, & Milberg, 1996; Revonsuo, 1995; Revonsuo & Laine, 1996), suggesting that there may be opportunities for successful intervention that capitalize on intact implicit abilities. These observations, combined with the logical inconsistency of using largely explicit therapy approaches to treat largely implicit language processes, suggest that there may be some benefit to exploring if anomia treatment could be improved through the use of largely implicit approaches to treatment, perhaps in conjunction with traditional explicitly based approaches.

^aUniversity of Washington, Seattle

Correspondence to JoAnn P. Silkes: jsilkes@uw.edu

Editor: Anastasia Raymer

Associate Editor: Heather Wright

Received September 12, 2014

Revision received February 10, 2015

Accepted May 11, 2015

DOI: 10.1044/2015_AJSLP-14-0138

Masked Priming Effects and Implicit Processing

One method that has been developed to measure and facilitate implicit processes while largely bypassing explicit, top-down influences is masked priming. Masked priming

Disclosure: The author has declared that no competing interests existed at the time of publication.

involves presenting prime words very briefly with competing visual stimuli presented before and after the prime. This combination interferes with conscious perception of the masked word (Forster, Mohan, & Hector, 2003) so that any processing of the prime is mediated solely through automatic, implicit processes, such as automatic spreading activation (Kiefer, 2002). We presume that masked primes activate their lexical-semantic representations so that these representations are more easily accessed when the target is presented. In the specific case of repetition priming, in which the prime and target are the same lexical item, priming effects in picture naming with visible primes are thought to occur due to increased efficiency of the object recognition and lexical retrieval processes in response to the prime (Francis, 2014). If automatic spreading activation—thought to support all levels of recognition and lexical retrieval—is activated through masked primes, then masked repetition primes should also show these effects.

Masked Repetition Priming in Aphasia Treatment

Indeed, a few prior studies have shown that using masked repetition primes can have a positive effect on naming in anomia. One study demonstrated greater response accuracy in an individual with anomia for words primed implicitly with a single presentation of masked prime words than for words that were unprimed (Avila, Lambon Ralph, Parcet, Geffner, & Gonzalez-Darder, 2001). These authors argued that automatic spreading activation was the source of the priming effects. Operating on the commonly accepted assumption that anomia involves intact semantic representations and processing but impaired access to word representations, they suggested that the masked primes served to support activation of the phonological forms of the target items, which then supplemented activation available from the semantic system, making the words more easily accessed when the target pictures were presented.

Subsequent studies have explored the possibility of extending the immediate masked priming effects demonstrated by Avila et al. (2001) by providing repeated pairings of masked prime words with target pictures across multiple training sessions with the objective of strengthening the connections between semantics and phonology that are needed to improve word retrieval. In two prior Phase 1 case studies, repeated exposure to masked identity primes paired with pictures has yielded improved naming of those pictures as well as improvements in various measures of broader language function (Silkes, Brookshire, Dierkes, & Kendall, 2012; Silkes, Dierkes, & Kendall, 2012). These studies each explored cumulative effects of multiple prime-target presentations and naming attempts per session over repeated sessions and provided initial encouragement that masked primes may have a useful role to play in rehabilitation of anomia. They were limited, however, by small numbers (a single case study per report) and by lack of adequate measures of within- and across-category generalization. In addition, there were some unexpected outcomes, including evidence for cross-category generalization and

improvements in broader language skills beyond naming—neither of which were predicted and, therefore, bear further exploration.

The present study, therefore, continues this line of investigation, extending masked repetition priming treatment for anomia to more participants and providing more thorough measures of generalization. Specifically, we asked if repeated exposure to masked repetition primes paired with pictures leads to the following:

1. Improved naming of those pictures. We hypothesized that trained items would improve due to strengthening of lexical networks through Hebbian learning (Hebb, 1949).
2. Improved naming of untrained pictures in the same category. We hypothesized that untrained items in the same semantic category would also improve due to overall strengthening of lexical networks surrounding trained items through automatic spreading activation.
3. Improved naming of pictures in other semantic categories. We hypothesized that there would be no cross-category generalization as unrelated lexical networks do not share many connections. Previous work, however, suggests that cross-category generalization could occur (Silkes, Dierkes, & Kendall, 2012); if this were found in this study, it would suggest overall strengthening of broad lexical retrieval mechanisms.
4. Improved general language function as measured by tests of both broader language function and discourse analysis. On the basis of previous work (Silkes, Dierkes, & Kendall, 2012), we predicted that participants would show improved scores on the Western Aphasia Battery (WAB; Kertesz, 1982) and Boston Naming Test (BNT; Kaplan, Goodglass, & Weintraub, 2001) as well as positive changes in connected discourse. If improved general language function were observed, it would be further evidence for strengthening of either broader lexical retrieval mechanism.

Method

Study Design

This early Phase 2 (Robey, 2004) study was a single-subject, multiple-baseline design, replicated across four participants with aphasia. Participants who completed the entire experiment received \$300.

Participants

Participants were recruited through the University of Washington Aphasia Registry and the UW Speech and Hearing Clinic. All participants had aphasia with anomia as verified by initial administration of the WAB and BNT. They also were tested with Raven's Coloured Progressive

Matrices (RCPM; Raven, 1976) to verify that there were no impairments reflecting potential right hemisphere lesions and with the first three subtests of the Reading Comprehension Battery for Aphasia (LaPointe & Horner, 1979) to verify intact single-word reading (see Table 1 for a summary of all participant initial test profiles). None had a significant history of psychiatric disorders, evidence of depression (as measured by a score < 4 on the Depression Intensity Scale Circles; Turner-Stokes, Kalmus, Hirani, & Clegg, 2005), or other neurological disease or injury beyond that which caused aphasia. All showed adequate visual acuity by passing a vision screening for 20/40 vision using a tumbling E eye chart and had no evidence of neglect or field cut as seen on line bisection.

Four out of six potential participants qualified for inclusion. One participant who was initially assessed did not qualify on the basis of scores on the RCPM, and another did not qualify because her naming performance on baseline probes was too high.

Setting and Equipment

Stimuli were delivered on a Dell desktop computer running Windows 7 with a 20-in. CRT computer screen. The refresh rate for the screen was set to either 70 or 100 Hz, depending on the prime exposure duration used for each participant to ensure millisecond accuracy of stimulus delivery. Experimental protocols were presented using E-Prime Professional 2.0 (version 2.0.10.242, Psychology Software Tools, Pittsburgh, PA). Verbal naming responses were digitally recorded (Olympus Digital Voice Recorder VN-4100PC).

Stimuli

Eleven semantic categories were created before beginning the experiment, to provide a broad range of options for participants. Items in each semantic category were chosen by the author. Categories were necessarily large and diverse given the number of items needed for each participant in this study design. Although some items in each category are more or less typical or common, all were determined by the author to be logically categorized in that group, and this was confirmed informally through discussion

with other members of the University of Washington Aphasia Research Laboratory. An additional semantic category was developed later to meet the needs of one participant who did not have enough items consistently named incorrectly in any other category as described below.

Picture stimuli for the treatment program were color photographs or drawings. All pictures were modified to remove relevant written information although a few items necessarily had some printed information still visible (e.g., letter tiles on a Scrabble board and the names of properties on a Monopoly board). Treatment stimuli were customized for each participant (see Appendix); only items familiar to the participant were used in his or her treatment program. Prime words were identity primes (i.e., the name of the upcoming picture) and were presented in 30-point, black, Arial font in the center of a computer screen.

Word frequencies were obtained from the Subtlex database (Brysbaert & New, 2009). For multiword items and proper names, which are not reflected in this database, values were calculated on the basis of the Corpus of Contemporary American English (Davies, 2008). Measures of phonotactic probability were obtained through an online probability calculator (Vitevitch & Luce, 2004).

Procedures

Baseline Naming Probes

Participants were seen for initial speech-language assessment to verify that they qualified for the study and to obtain initial baseline measures of language function. Before baseline naming probes were conducted, participants chose three semantic categories that they thought would be particularly difficult for them with the intention of identifying two that could be included in the treatment protocol (i.e., that had enough items that they could not reliably name across baseline probes). Baseline testing sessions with these three categories were conducted across 4 days with at least a 1-hr break between baselines conducted on a single day. During these sessions, participants were seated in front of a computer and attempted to name pictures presented one at a time for up to 10 s in the center of the computer screen. If they were unable to name a picture, they were asked if they recognized it and asked for information

Table 1. Participant profiles.

Participant	Age	Medical Dx	Time postonset	SLP Dx	RCPM (/36)	RCBA (/30)	Initial WAB AQ (/100)	Initial BNT (/60)
P1	60	Left MCA CVA	8 years	Moderate nonfluent aphasia with mild-to-moderate apraxia of speech	33	30	69.1	15
P2	62	Left MCA CVA	2 years	Moderate fluent aphasia	35	29	73.9	30
P3	60	Left MCA CVA	5 years	Mild fluent aphasia	32	29	89	40
P4	59	SAH from PCA aneurysm	4.5 years	Moderate fluent aphasia	29	27	68.5	11

Note. Dx = diagnosis; SLP = speech-language pathology; RCPM = Raven's Coloured Progressive Matrices; RCBA = Reading Comprehension Battery for Aphasia; WAB = Western Aphasia Battery; BNT = Boston Naming Test; MCA = middle cerebral artery; CVA = cardiovascular accident; SAH = subarachnoid hemorrhage; PCA = posterior cerebral artery.

about it to verify that they knew what it was. Stimuli were randomly presented within each semantic category list with a break provided between lists. The order of category presentation was rotated for each session. If baseline testing revealed that any of the categories selected were too easy for the participant (i.e., they did not yield enough items for treatment in at least two categories), additional categories were tested in baseline naming probes until two categories were identified for inclusion in the treatment portion of the study. Because of this need to adjust the categories being tested as baseline probes proceeded, semantic categories were not mixed for the baseline naming probes. This practice of blocking naming probes by semantic category was then maintained for all naming probes throughout the remainder of the protocol. A total of seven baseline naming probes were completed for all categories selected for training.

In addition to the baseline naming probes, conversational discourse samples were analyzed for percentage of Correct Information Units (CIUs; Nicholas & Brookshire, 1993) before and after treatment for all but one participant. This measure was incorporated to explore whether improved lexical access from the structured training task would be reflected in improved use of specific, relevant vocabulary in connected speech. This task involved recording participants as they spoke freely in response to open-ended questions (“Tell me about what kinds of medical problems you’ve been dealing with;” “Tell me about what you do in a typical day from when you wake up in the morning until you go to bed at night”) with additional prompts and questions asked as needed to elicit further responses.

Last, the Five Point Test (Regard, Strauss, & Knapp, 1982) was administered repeatedly across the entire protocol as a control measure of broader changes in nonlinguistic cognitive abilities. This task involves the participant creating as many different designs as possible by connecting dots in a 5-dot array (: • :) without lifting the pencil from the paper or retracing lines. The scores represent the percentage of unique (i.e., nonrepeated) designs produced in a 3-min period. As a nonlinguistic task that reflects nonverbal organization and planning, no systematic change is expected in this task from before to after treatment.

Stimulus Selection

Once two semantic categories were identified, only items that were named correctly three or fewer times, and not in the final baseline naming probe, were considered for inclusion. Of these, 24–30 were selected and distributed across three conditions: trained (T), untrained–exposed (UE), and untrained–unexposed (UU). Trained items were seen throughout the treatment protocol accompanied by masked identity primes. UE items were seen an equal number of times as the trained items but were accompanied by masked sham primes (strings of alternating *xs* and *gs* that equaled the length of the target word). UU items were seen only during naming probes but not during treatment sessions. Lists were balanced for word frequency, number of letters, number of phonemes, and phonotactic probability. In addition, an attempt was made to distribute items that

were particularly closely related to each other across the three different conditions (e.g., *Operation*, *Monopoly*, and *Scrabble* as table games within the larger category of sports and games; *necklace*, *pin*, and *earring* as jewelry within the larger category of things to wear). This distribution was intended to make sure that there were at least some closely related items across the conditions, facilitating the chances for observing within-category generalization due to shared activation, and to equalize relatedness within conditions. For most participants, all conditions comprised the same number of items (one exception is described below).

Prime Visibility Threshold Testing

Prior to beginning treatment sessions, participants were also tested to determine the exposure duration at which prime items should be presented so as to minimize task-relevant conscious awareness of them (Reingold, 2004). A category judgment task was used, in which participants were asked to make judgments on sets of masked and unmasked words. In each trial, a masked word or non-word was presented, followed by a clearly visible word (see Figure 1). Participants pressed a button if they saw a word that was something to eat or drink. Practice trials were completed with 300- and 100-ms exposure durations for the masked items, and the exposure duration was then reduced over successive lists, testing performance at 30, 20, and 10 ms. An additional exposure duration of 14 ms was added beginning with the second participant. The longest exposure duration at which masked primes were accurately categorized (i.e., category member or not) on <60% of the trials was selected as that participant’s prime exposure duration for the experiment.

Treatment Session Protocol

In each treatment session, the participants were seated at a comfortable distance from the computer screen. Treatment lists included both T and UE items, each presented one time in random order in each of four training runs. Each training list contained all T and UE items, presented in random order. A training session involved presentation of four lists with breaks between lists as long or short as the participant desired (most were less than 1 min). A single trial involved four presentations of the masked prime-target picture pair with 1 s separating each presentation (details of the stimulus presentation sequence are provided in Figure 2).

Figure 1. Schematic representation of visibility threshold testing task. Participants pressed a button if they saw a food or drink word in either the masked or visible position.

Therefore, in a single session, the participant saw each prime-picture a total of 16 times (four presentations in each of four lists) with four opportunities to name each picture (on the fourth presentation within each list). No mention was made of the masked primes; participants were told to just watch everything that came up on the screen and then try to name the picture when they saw it for the fourth time. On this fourth presentation, the picture was bordered in green, signaling participants that it was time to respond. If they did not name it before the target picture presentation ended after 10 s, they were encouraged to just let it go and prepare for the next item to begin 4 s after the target picture disappeared. No feedback was given at any time.

In a typical treatment day, two treatment sessions were conducted with a 1-hr break between sessions. Twelve treatment sessions (6 days) were conducted for each semantic category. Treatment naming probes were conducted prior to every third treatment session, at the beginning of each treatment day starting on the second day. Both T and UE items were seen during all naming probes. For the first three participants, all naming probes also included the UU items. For the fourth participant (P4), UU items were shown only during baseline, the final three posttraining, and maintenance probes but not during training probes to truly minimize exposure to those items. Three posttreatment naming probes and follow-up language testing were conducted over 2 days in the week immediately following completion of training for the second semantic category. Three additional maintenance naming probes and repeated language testing were conducted approximately 3 months after that. All treatment sessions and naming probes were audio-recorded for later analysis.

Data Processing and Analysis

Response accuracy was noted by the experimenter during all naming probes and treatment sessions with questionable responses verified from the audio recordings before being scored. Naming responses were considered accurate if the participant produced the target name of the picture at any time during the 10-s response window. Words with sound distortions were counted as correct only if the distortion did not cross phonemic boundaries. Responses that varied from the target by inflectional morphemes were counted as correct (e.g., *carrots* for *carrot*), but responses

that involved a derivational morpheme change were counted as incorrect (e.g., *juggle* for *juggling*). Words that were pronounced with all phonemes correct but had incorrect prosody (e.g., *ro-DE-o* for *RO-de-o*) were counted as correct as long as they were recognizable.

Naming probe data for all participants were analyzed by calculating effect sizes according to the procedure described by Busk and Serlin (1992): $[ES = (Mean_{posttreatment} - Mean_{baseline}) / SD_{baseline}]$. For the first semantic category trained for each participant (L1), this was done comparing mean naming accuracy on the eight naming probes conducted during and immediately following treatment of the second semantic category (L2) with mean accuracy on the seven pretreatment naming probes relative to baseline standard deviations. For L2, pretreatment naming probes included means and standard deviations from all extended baseline probes, including those conducted before and during treatment of L1, and posttraining probes included the three probes administered immediately following completion of treatment. Maintenance probes included the three probes administered 3 months following completion of the posttraining probes. Cross-category generalization was calculated by comparing the eight naming probes administered for L2 during training of L1 with the seven L2 baseline probes. Effect sizes were interpreted relative to the criteria determined by Beeson and Robey (2006) with effects >2.6 considered small, >3.9 considered medium, and >5.8 considered large.

Reliability

A minimum of 25% of all naming probes for each participant were scored a second time by research assistants listening to the audio recordings. Scoring reliability for these probes was calculated using Cohen's kappa, comparing the author's scoring with the research assistants' scoring. Rater agreement was *substantial* to *almost perfect* (Landis & Koch, 1977) with scores of .883 (P1), .944 (P2), .989 (P3), and .776 (P4). Any disagreement between judges was resolved by the author before analysis on the basis of audio recordings and written notes (e.g., if a correct response was made after the response deadline, it would be marked as incorrect on the response sheet with a note indicating that it was late, but the research assistant may have counted it as correct).

Figure 2. Schematic representation of a single trial of the treatment task. Note that the variable exposure duration for masked primes was set individually for each participant as described in the text.

For language samples, a research assistant transcribed all samples from audio recordings. The author then reviewed all transcripts while listening to the recordings to verify their accuracy. Any questionable utterances were reviewed by members of the research team to achieve consensus on how they should be transcribed.

Results

Participant 1

Participant 1 (P1) was a 60-year-old man who was 8 years postonset of a large left hemisphere middle cerebral artery (MCA) stroke, encompassing the frontal, temporal, and parietal lobes. Mild-to-moderate apraxia of speech was noted with a slowed rate of speech, occasional sound distortions, and abnormal prosody. Frequent hesitations were noted in connected speech with occasional semantic paraphasias in the context of verbal output that was generally short phrases to short sentences. He had good error recognition and occasional single-word to short-phrase repetitions as he attempted to correct errors in connected speech. He often self-cued for word retrieval by reciting a portion of the alphabet leading up to the first letter of the target word (e.g., *m, n, o, p ... pear*). His prime exposure duration was set at 10 ms as per visibility threshold testing. He was seen for only two sessions per week due to the distance he had to travel to attend sessions.

Stimuli and Stimulus Selection

The categories of produce (L1) and sports and games (L2) were selected for treatment (see Appendix, Table 1). Each category comprised 30 items.

Results

During treatment sessions, naming accuracy for primed items relative to unprimed items approached significance for L1, $t = 1.89$, $p = .09$, and was significant for L2, $t = 2.66$, $p = .022$. Effect sizes for picture naming probes immediately posttreatment were 2.94 (small) for T, 5.28 (medium) for UE, and 0.61 (ns) for UU items in L1 and 3.87 (small) for T, 3.08 (small) for UE, and 1.74 (ns) for UU items in L2 (see Figure 3 for all data from P1 and Table 2 for all participants' effect sizes). He showed an effect size of 4.36 (medium) for cross-category generalization. At maintenance, effect sizes were 2.94 (small) for T, 3.56 (small) for UE, and 1.39 (ns) for UU items in L1 and 3.48 (small) for T, 2.68 (small) for UE, and 1.3 (ns) for UU items in L2.

Because cross-category generalization was evident for L2 (i.e., naming probe accuracy for L2 increased during training of L1), pre- to posttreatment comparisons that include the extended baseline for L2 confound the ability to identify the effect of the training on naming probe items. Therefore, effect sizes for L2 were recalculated (for all participants) using only the first seven baseline probes, conducted before any training took place at all. These baselines were compared with the immediate posttraining probes as described above. In this comparison, an effect size of

10.24 (large) was noted for T, 8.44 (large) for UE, and 3.51 (small) for UU items. Recalculating the maintenance data using only the first seven baselines for L1 yielded an effect size of 9.32 (large) for T, 7.56 (large) for UE, and 2.83 (small) for UU items.

During baseline naming probes, errors were often neologistic, usually showing partial retrieval of phonological forms (*zuzickers* for *zucchini*, *va ... vole ... volri ... vack ... vollery* for *volleyball*). Other errors included semantic similarity to targets (*horsebacking* for *rodeo*) or involved self-cuing with the alphabet (*a-b-c-core ... corn ... canner* for *canoeing*), and he showed very occasional perseverations from previous probe items. Perseverations were not widely noted during treatment sessions. During posttreatment naming probes and at maintenance, errors tended to be partial or incorrect grammatical forms of the correct response (*canoe* for *canoeing* or *bull riding* for *bullfighting*), phonemic paraphasias (*artery* for *archery*, *leatherball* for *tetherball*), or descriptions that he settled on during treatment and maintained reliably over time (*exercise mat* for *gymnastics*). Some neologistic responses persisted, but they were more consistent than during the baseline period (e.g., consistent production of *asperpyers* for *asparagus*) with fewer varying productions and efforts at self-correction (as shown for *volleyball*, above) noted.

WAB-Phasia Quotient (WAB-AQ) scores before, immediately after, and 3 months after treatment were 69.1, 69.9, and 71.2, respectively. BNT scores were 15, 17, and 29, respectively (see Tables 3, 4, and 5 for a summary of all participants' test data). Discourse data were not obtained for this participant. Comparison of the percentage of accurate and unique designs produced on the Five Point Test during the final three posttraining sessions with baseline yielded an effect size of -1.67 (ns); at the time of maintenance testing, the effect size was 0.84 (ns).

Participant 2

Participant 2 (P2) was a 62-year-old man who was 2 years postonset of a left hemisphere MCA stroke. He presented with moderate fluent aphasia. No dysarthria or apraxia of speech was noted. Verbal output was generally fluent and grammatic with frequent hesitations, circumlocution, and single-word and short-phrase repetitions in response to word retrieval failures. Speech errors in connected speech were a combination of semantic and phonologic errors with frequent use of nonspecific vocabulary. During training sessions, he demonstrated frequent perseverations on whole words across training items.

Stimuli and Stimulus Selection

The categories occupations (L1) and vehicles (L2) were selected for training (see Appendix, Table 2). He did not have 30 items in each category that were consistently incorrect, however, so training proceeded with 22 occupations and 25 vehicles that he had named correctly no more than three times during baseline testing. They were then assigned to one of three conditions for the training portion

Figure 3. Data from participant P1.

of the study with 10 T, 6 UE, and 6 UU items in L1 and 10 T, 10 UE, and 5 UU items in L2. His prime exposure duration was 14 ms. He was seen 4 days per week.

Results

During treatment sessions, primed items were named significantly more accurately than unprimed items in L1, $t = 3.83$, $p = .003$, but unprimed items were more accurate than primed in L2, $t = -4.34$, $p = .001$. Immediately following training, effect sizes for picture naming probes were 3.19 (small) for T, 0.31 (ns) for UE, and 0.99 (ns) for UU items in L1 and 1.66 (ns) for T, 0.63 (ns) for UE, and 0.88 (ns) for UU items in L2 (see Figure 4 for all data from P2). He showed an effect size of 0.62 (ns) for cross-category generalization. Recalculating effect sizes for L2, using only the first seven baseline probes as described above, he showed an effect size of 2.02 (small) for T, 0.71 (ns) for UE, and 1.27 (ns) for UU items. At maintenance, effect sizes were 0.82 (ns) for T, -0.82 (ns) for UE, and 1.64 (ns) for UU items in L1 and 1.07 (ns) for T, 0.13 (ns) for UE,

and -0.92 (ns) for UU items in L2. Recalculating maintenance data from L2 using only the first seven baseline probes yielded an effect size of 1.44 (ns) for T, 0.24 (ns) for UE, and -0.78 (ns) for UU items.

During baseline naming probes, errors were primarily semantically related (*baseball* or *basketball* for *umpire*) or descriptive (*he's at war* for *soldier*) with occasional phonemic paraphasias (*pharmastist* for *pharmacist*) and occasional perseverations, which were typically on related items (e.g., accurately naming a *car* but then also using *car* for *bus* a few items later). Error types were largely the same during posttreatment naming probes and at maintenance testing although he tended to produce more facilitative carrier phrases than previously (*he's the head coach in football* for *coach*), most of which he had developed as self-cues during treatment sessions.

WAB scores before, immediately posttreatment, and at maintenance testing were 73.9, 78.5, and 74.6, respectively. BNT scores were 30, 31, and 26. In connected speech, he produced 61% CIUs at initial testing and 69% CIUs at

Table 2. Summary of effect sizes (*d*) for all participants.

Condition	L2 calculated using extended baseline probes							
	P1		P2		P3		P4	
Immediately posttreatment	L1	L2	L1	L2	L1	L2	L1	L2
Trained	2.94 ^a	3.87 ^a	3.19 ^a	2.14	18.05 ^c	7.88 ^c	4.81 ^b	3.67 ^a
Untrained–exposed	5.28 ^b	3.08 ^a	0.15	0.47	4.15 ^a	4.27 ^b	1.77	–0.76
Untrained–unexposed	0.61	1.74	0.99	0.83	2.29	1.75	5.72 ^b	
Cross-category generalization		4.36 ^b		0.62		6.24 ^c		1.39
Maintenance								
Trained	2.94 ^a	3.48 ^a	0.82	1.43	4.15 ^b	4.85 ^b	4.18 ^b	1.62
Untrained–exposed	3.56 ^a	2.68 ^a	–0.82	0.21	2.60 ^a	2.89 ^a	1.07	1.22
Untrained–unexposed	1.39	1.3	1.64	–1.19	0.78	2.51	1.64	
L2 recalculated using only the first seven baseline probes								
Immediately posttreatment								
Trained		10.24 ^c		2.75 ^a		13.89 ^c		3.09 ^a
Untrained–exposed		8.44 ^c		0.47		11.97 ^c		–0.38
Untrained–unexposed		3.51 ^a		1.25		2.45		2.27
Maintenance								
Trained		9.32 ^c		2.01		8.99 ^c		1.39
Untrained–exposed		7.56 ^c		0.24		8.44 ^c		2.27
Untrained–unexposed		2.83 ^a		–1.39		3.27 ^a		0.50

Note. As described in the text, untrained–unexposed items were not assessed during extended baselines for P4.

^asmall effect. ^bmedium effect. ^clarge effect.

3-month maintenance testing; this measure was not taken immediately posttreatment due to time constraints. Comparison of the percentage of accurate and unique designs produced on the Five Point Test during the final three posttraining sessions with baseline yielded an effect size of 0.22 (ns); at the time of maintenance testing, the effect size was 0.12 (ns).

Participant 3

Participant 3 (P3) was a 60-year-old man 5 years post left MCA cerebrovascular accident. He presented with mild fluent aphasia. Connected speech was typically fluent and grammatic with frequent hesitations, interjections (*um*), and occasional word or phrase repetitions, all in the face of word retrieval failures. He often self-cued for word retrieval by writing in the air with his finger with varying degrees of success. Word retrieval errors manifested primarily in partial productions or circumlocutions rather than producing overt errors although occasional semantic paraphasias were noted. When errors occurred, error recognition was good. No dysarthria was noted, but he showed evidence of minimal-to-mild apraxia of speech.

Stimuli and Stimulus Selection

When P3 did not meet inclusion criteria on baseline testing of any of the 11 available semantic categories, two additional categories were created on the basis of his suggestions about what words were particularly difficult for him in daily life: famous faces (FF) and brands and logos. In baseline naming probes with these categories, FF was identified as appropriate for treatment with adequate numbers of people he could not name consistently to create

two subgroups. L1 was labeled FF-entertainment and comprised 30 targets that included actors, musicians, and comedians. L2 was labeled FF-other and comprised 24 targets that included people from venues such as politics, news, and sports (see Appendix, Table 3). Although creating subgroups from a single broader category may limit the ability to distinguish within- versus cross-category generalization effects, it was determined that valuable data would be obtained despite this limitation given that this participant would not have been able to complete the experiment otherwise. His prime exposure duration was set at 14 ms. He was seen 4 days per week.

Results

During treatment sessions, primed items were named significantly more accurately than unprimed items in L1, $t = 8.92$, $p < .000$, but there was no difference in accuracy for primed and unprimed items in L2, $t = 1.35$, ns. Immediately following training, effect sizes for picture naming probes were 18.05 (large) for T, 4.15 (small) for UE, and 2.29 (ns) for UU items in L1 and 7.88 (large) for T, 4.27 (medium) for UE, and 1.75 (ns) for UU items in L2 (see Figure 5 for all data from P3). He showed an effect size of 6.24 (large) for cross-category generalization. At maintenance, effect sizes were 4.15 (medium) for T, 2.6 (small) for UE, and 0.78 (ns) for UU items in L1 and 4.85 (medium) for T, 2.89 (small) for UE, and 2.51 (ns) for UU items in L2. After recalculating effect sizes for L2 using only the first seven baseline probes as described above, he showed an effect size of 13.89 (large) for T, 11.97 (large) for UE, and 2.45 (ns) for UU items. Recalculating maintenance data for L2 using only the first seven baselines

Table 3. Summary of pre-, post-, and maintenance Western Aphasia Battery–Aphasia Quotient (WAB AQ) and subtest scores.

WAB Task	Pre	Post	Maintenance
P1			
AQ	69.1	69.9	71.2
Spontaneous speech	9	9	15
Information content	5	5	9
Yes/no questions	57	57	57
Auditory word recognition	58	58	59
Sequential commands	70	66	66
Repetition	40	44	36
Object naming	48	48	52
Word fluency	9	9	9
Sentence completion	8	8	8
Responsive speech	8	10	10
P2			
AQ	73.9	78.5	74.6
Spontaneous speech	8	8	8
Information content	9	9	8
Yes/no questions	54	60	54
Auditory word recognition	49	47	49
Sequential commands	50	80	37
Repetition	71	67	75
Object naming	37	44	43
Word fluency	1	2	7
Sentence completion	8	6	10
Responsive speech	8	10	8
P3			
AQ	89	85.2	86.9
Spontaneous speech	10	9	9
Information content	9	9	9
Yes/no questions	54	57	54
Auditory word recognition	59	57	57
Sequential commands	57	58	50
Repetition	77	73	77
Object naming	56	54	57
Word fluency	17	13	20
Sentence completion	10	10	10
Responsive speech	10	10	10
P4			
AQ	68.5	71.7	71.6
Spontaneous speech	9	9	9
Information content	9	9	8
Yes/no questions	48	51	51
Auditory word recognition	45	46	51
Sequential commands	18	8	38
Repetition	58	62	58
Object naming	31	38	32
Word fluency	6	10	10
Sentence completion	6	6	8
Responsive speech	6	10	10

yielded effect sizes of 8.99 (large) for T, 8.44 (large) for UE, and 3.27 (small) for UU items.

During baseline naming probes, errors generally involved descriptions (*Russian* for *Mikhail Gorbachev*), parts of names (e.g., first or last name only), occasional phonemically related attempts (*Crosby* for *Bill Cosby*), and omissions (e.g., *I don't know* or no response). During posttreatment naming probes and at maintenance testing, fewer descriptions were noted, but omissions and phonemically related attempts were still prevalent.

WAB scores before, immediately after, and 3 months after treatment were 89, 85.2, and 86.9, respectively. BNT

Table 4. Summary of pre-, post-, and maintenance Boston Naming Test scores for all participants.

Participant	Pre-tx	Post-tx	Maintenance
P1	15	17	29
P2	30	31	26
P3	40	45	46
P4	11	13	8

scores were 40, 45, and 46. In connected speech, he produced 73% CIUs at initial testing and immediately following treatment and 78% CIUs at 3-month maintenance testing. Comparison of the percentage of accurate and unique designs produced on the Five Point Test during the final three posttraining sessions with baseline yielded an effect size of 0.70 (ns); at the time of maintenance testing, the effect size was -0.55 (ns).

Participant 4

Participant 4 (P4) was a 59-year-old woman 4.5 years post subarachnoid hemorrhage due to a posterior cerebral artery aneurysm. Verbal output in connected speech was fluent and grammatic with primarily phonologic paraphasias, occasional semantic paraphasias, and frequent use of nonspecific vocabulary and circumlocution that tended to lack detail. Occasional phrase repetitions were noted when she attempted to repair word retrieval failures. Her awareness of overt production errors was good in spontaneous speech, but she generally did not recognize when she had not provided specific information or vocabulary. On confrontation naming tasks, she did not tend to benefit from phonemic cues and occasionally would spell aloud words that she could not produce. Her naming performance tended to decrease as tasks progressed, with increasing perseveration, and she benefitted from frequent breaks. During training sessions, she frequently perseverated on whole words across training items. Perseveration was also frequently seen on her responses on the Five Point Test. No dysarthria or apraxia of speech were noted. Although she met inclusion criteria on the RCPM, she demonstrated some behavioral evidence of potential impairments in long- and short-term memory in daily interactions.

Stimuli and Stimulus Selection

The categories things to wear (L1; 27 items) and sports and games (L2; 30 items) were selected for treatment (see Appendix, Table 4). P4's prime exposure duration was determined to be 20 ms. Due to an E-Prime programming error, however, only the first presentation of each prime in a trial was 20 ms long, and the remaining three were 14 ms long. She was seen four times per week.

Results

During treatment sessions, primed items were named significantly more accurately than unprimed items in both L1,

Table 5. Summary of pre-, post-, and maintenance percentage Correct Information Unit (CIU) counts for all participants.

Participant	Pre			Post			Maintenance		
	Total words	Total CIUs	% CIUs	Total words	Total CIUs	% CIUs	Total words	Total CIUs	% CIUs
P1	—	—	—	—	—	—	—	—	—
P2	678	416	61%	—	—	—	1001	686	69%
P3	767	560	73%	866	629	73%	445	349	78%
P4	849	549	65%	1030	690	67%	963	716	74%

Note. Em dashes indicate data not available.

$t = 3.81$, $p = .003$, and L2, $t = 4.35$, $p = .001$. Immediately following training, effect sizes for picture naming probes were 4.81 (medium) for T, 1.77 (ns) for UE, and 5.72 (medium) for UU items¹ in L1, and 3.67 (small) for T and -0.76 (ns) for UE items² in L2 (see Figure 6 for all data from P4). She showed an effect size of 1.39 (ns) for cross-category generalization. At maintenance, effect sizes were 4.18 (medium) for T, 1.07 (ns) for UE, and 1.64 (ns) for UU items in L1 and 1.62 (ns) for T and 1.22 (ns) for UE items in L2. After recalculating effect sizes for L2 using only the first seven baseline probes, she showed an effect size of 3.09 (small) for T, -0.38 (ns) for UE, and 2.27 (ns) for UU items immediately posttreatment, and 1.39 (ns) for T, 2.27 (ns) for UE, and 0.50 (ns) for UU items at maintenance.

During baseline naming probes, errors were typically omissions (*I don't know*, *I know that one*, or no response), semantic paraphasias (*dress* for *apron*), phonemic paraphasias (*scotter* for *soccer*), or descriptions (*cowboy on the horse* for *rodeo*). During posttreatment naming probes and at maintenance testing, errors were largely similar with the exception of having few omissions and more frequent perseverations. These perseverations were typically—although not always—within-category and were often consistent with perseverative errors that she had made during treatment sessions (e.g., *white top* for *blouse*, *short top* for *shirt*, *long top* for *sweatshirt*, *hanger top* for *juggling*, followed by *ball hanger* for *tetherball* and *kickball*, *hanger top* for *diving*, and *ankle top* for *javelin*).

WAB scores before, immediately following, and 3 months following treatment were 68.5, 71.7, and 71.6, respectively. BNT scores were 11, 13, and 8. In connected speech, she produced 65% CIUs at initial testing, 67% immediately following treatment, and 74% CIUs at 3-month maintenance testing. Comparison of the percentage of accurate and unique designs produced on the Five Point Test during the final three posttraining sessions with baseline yielded an effect size of 1.41 (ns); at the time of maintenance testing, the effect size was 1.07 (ns).

¹The effect size for UU items in L1 was calculated using just the last three posttraining probes as these items were not seen during any training probes.

²An effect size for UU items in L2 could not be calculated with an extended baseline because these items were not seen during any training probes.

Discussion

The four cases presented here represent the findings of a preliminary investigation of the potential for using masked repetition priming as a tool for improving word retrieval in aphasia with anomia. In general, repeated exposure to masked primes paired with pictures (T condition) yielded better improvements in naming than repeated exposure to the pictures alone (UE condition); indeed, for P2 and P4, changes were seen only for T items and not for UE items. For P1 and P3, who showed improvement of both T and UE items (with T items improved more than UE items in all but L1 for P1), both of these conditions were better than no exposure during treatment (UU condition). Contrary to the prediction outlined earlier, there was no consistent improvement in naming of items in the UU condition, indicating no consistent within-category generalization. Finally, gains were maintained for both trained categories for two participants (P1 and P3) and for one category for one participant (P4). Nonsignificant outcomes on the nonlinguistic control measure, the Five Point Test, demonstrate that any changes observed were not likely related to overall improvements in cognitive abilities.

These findings suggest that repeated exposure alone had some positive lasting effects on naming success, consistent with other available evidence (Off & Griffin, 2014), but that adding a masked prime to the exposure sequence generally created a benefit beyond that of just repeated exposure. These findings are also consistent with previously published pilot data (Silkes, Dierkes, & Kendall, 2012). In contrast with the pilot data, however, these participants did not make notable improvements in broader language measures of the WAB or BNT although small gains were noted in the percentage of CIUs in connected speech.

Although a complete error analysis was not conducted, some patterns were evident across participants in terms of the errors produced before versus after treatment. All participants became more consistent in their naming responses across the course of the experiment, often even with their incorrect responses. In many cases, during treatment sessions, participants would settle on a name for a picture that was not correct but was then carried forward through the remainder of the protocol (e.g., P1's consistent production of *sub attack* for *Battleship* and *asperpyers* for *asparagus* and P4's relatively consistent production of *eyeglass holder* for *goggles* and *red hanger top* for *suspenders*). Participants

Figure 4. Data from participant P2.

often became confident in these responses and did not appear to make an effort to correct them further. This observation raises the question of the role of feedback in a protocol such as this one. Although some prior literature suggests that learning can occur in anomia treatment without feedback (Fillingham, Sage, & Lambon Ralph, 2005b; Off & Griffin, 2014), seeing these participants adopt and maintain incorrect responses in this protocol suggests that there may be a role for some form of feedback to maximize improvements with a masked priming protocol. If this feedback would best be knowledge of performance or knowledge of results (Schmidt & Lee, 2005) remains to be tested.

It is not surprising that there was substantial variability between participants. These differences may provide useful insights into the mechanisms by which masked primes may be effective and for whom this approach may be most appropriate. The patterns of response suggest that there are participant, stimulus, and protocol factors that may contribute to the noted variability.

The Influence of Participant Factors on Response to Treatment

It appears from these data that individual language profile, the presence of perseveration, and the role of explicit error awareness and recall may all influence participant response to a masked priming protocol. In terms of language profile, reading skills may be particularly critical for use of this type of prime. This seems like an obvious conclusion but stands in contrast to prior literature that suggests individuals with severe acquired alexia are nonetheless able to demonstrate intact single-word processing in implicit tasks (Mimura et al., 1996; Revonsuo, 1995). All of the participants had good single-word reading as measured by the Reading Comprehension Battery for Aphasia, but P3, who had the best response, had the strongest functional reading skills of the four participants (i.e., he regularly read newspaper articles and books, and although he tended to read slowly, he generally demonstrated good comprehension). P2, who had the poorest

Figure 5. Data from participant P3.

response, however, had the poorest functional reading, with evidence of deep dyslexia characterized by frequent semantic errors in reading connected text and poor processing of nonwords. Although it would seem that good single-word reading would be adequate for ensuring the ability to use the single-word primes, it is possible that poorer reading of longer material reflects reduced automaticity of word recognition, making it more difficult to take advantage of the masked prime words. This interpretation is consistent with prior evidence that different degrees of activation across varied levels of severity of reading impairment may account for differences in the ability to make use of masked primes (Roberts, Lambon Ralph, & Woollams, 2010). At the same time, however, P2 demonstrated priming effects to masked identity primes (as measured by response times on a supplemental lexical decision task) even though he showed the least response to this treatment approach. This finding suggests that his system may have been processing the primes but was unable to benefit from them in this particular task, implying that, even with repeated presentation, they were not activating lexical representations strongly enough to influence subsequent naming attempts. Further investigation of the

influence of reading ability on response to masked primes is warranted to better understand these issues.

The second, and related, participant factor to consider on the basis of these data is the presence of perseveration. Both P2 and P4, who were least responsive to this treatment, were highly perseverative during the treatment tasks and naming probe sessions, which appeared to influence their ability to self-correct in these contexts although perseveration was less evident in their discourse. Because of their perseverations, these participants may not have been able to take full advantage of the prime words within the timing parameters at which they were presented. If these participants have impaired timing of automatic spreading activation mechanisms as a cause of their perseveration, then previously activated items may not have decayed adequately at the time of presentation of the next stimulus (Martin & Dell, 2007), thus providing competition that was difficult to overcome. This is especially true for P4, who showed decreased performance over time in most tasks, reflecting a need to allow activation to dissipate before being successful again. It may be that these participants would have responded better to different timing parameters, such as longer intervals between trials to permit

Figure 6. Data from participant P4.

time for activation of the previous items to subside before introducing something new.

The third participant factor that these data suggest may be important is the degree of insight a participant has, including error recognition and explicit memory for prior productions. Both P1 and P3, who showed the greatest improvement for UE items and the best retention of naming improvements at maintenance testing, demonstrated excellent insight into their errors and memory for previous naming attempts. In both cases, they frequently appeared to recall previous responses to particular stimuli and to gradually build or retrieve the appropriate name, often piece by piece. In contrast, P2 and P4 showed poorer error awareness and did not often make responses that suggested they were building upon prior responses. These results suggest that, although the active stimuli for this task are implicit, there is still a significant role of explicit reflection and recall when attempting to achieve progressive improvement over multiple sessions. Consistent with prior literature on the role of executive and problem-solving skills in errorless learning in anomia treatment (Fillingham, Sage, & Lambon Ralph, 2005a), it seems that awareness and recall of errors may be critical components for successful

improvement of naming across sessions in the context of this treatment approach; participants may use this information to guide the responses they choose to produce, combining the effects of implicit primes with explicit reflection.

The Influence of Protocol Factors on Response to Treatment

Along with the issue of stimulus presentation timing, as discussed above, there is the question of how long prime words were presented, which is tied to participant awareness of the primes. Prime exposure durations were set individually in an effort to ensure that participants were not consciously aware of the content of the primes, and participants were not told that primes were present. Despite these procedures, some participants demonstrated awareness of something in the prime position and occasionally even showed some awareness of what it said. It is possible that differences between participants in the level of awareness of the primes contributed to differences in overall success of the protocol. It does not appear, however, that the prime words were fully visible to any of the participants.

Even for P3, who gave the greatest indication that he occasionally recognized the primes, accuracy for naming primed items during treatment sessions was variable. Given his high reading level, though, wholly visible primes would be expected to have led to immediate and consistent accurate naming of the targets. He may have occasionally truly seen the prime word, but it is also possible that his occasional feelings that he had consciously read the prime words resulted from retrospective interpretation or analysis of what he had seen after he had retrieved the target name. Being certain about prime visibility is important in this paradigm for understanding the locus of treatment effects; if it is unclear whether primes are consciously perceived or not, then priming effects cannot be confidently attributed to implicit rather than explicit processing. Future research is needed to further investigate the impact of prime awareness and variation of other prime characteristics, such as prime exposure duration, measures of visibility, and prime target intervals.

The second protocol factor that may influence participant performance is whether treatment stimuli are blocked by semantic category or mixed. In this study, treatment stimuli were blocked by semantic category in an effort to identify within- and cross-category generalization effects. The finding that cross-category generalization occurred for some participants in the absence of generalization to untrained items within the trained semantic category is surprising; on the basis of a spreading activation model of lexical processing, one would expect items within the same semantic category as trained items to benefit most from any increased network activation resulting from the masked primes. The lack of consistent within-category generalization may have been due to the broad definitions of the categories, leading to a limited spread of activation between items that were presumed to be related. At the same time, it should be noted that P4, the only participant who did not see the UU items repeatedly during all naming probes, was the only participant who showed significant within-category generalization effects (for one category). It may be worth considering if there was an advantage to not having the UU items presented repeatedly in the presence of the items that were practiced throughout the experimental protocol; this may have avoided within-category interference. This conjecture is supported by also finding the cross-category generalization for some participants, both in this study and in previous work (Silkes, Dierkes, & Kendall, 2012). We suggest that this pattern of generalization may be the effect of blocking stimuli by semantic category, creating competition and interference within categories (Schnur, Schwartz, Brecher, & Hodgson, 2006). Although same-category items that are frequently practiced together may suffer due to interference, items in the same category that are not frequently practiced will not experience this interference but will benefit from spreading activation from the trained items. In a similar manner, items in other categories may still show improvement if this treatment is leading to overall improved spreading activation within the lexical system. Future studies are needed to investigate the effects of manipulating semantic

blocking in this paradigm (e.g., comparing semantically blocked vs. mixed training lists).

The Influence of Stimulus Factors on Response to Treatment

Related to the question of prime exposure duration is if the presence of the prime items actually immediately improved naming of target items during the training sessions. This project is based on Avila et al.'s (2001) finding of improved naming in the presence of masked primes, but that work has not been replicated. Therefore, treatment data in this study were analyzed to determine whether masked priming led to better naming during treatment sessions. For most of the participants in most conditions, primed words were named accurately more often during treatment sessions than unprimed words, but this was not always the case. It is unclear why these differences across participants and semantic categories may have occurred. One possibility is that the effect of the masked primes was moderated by differences in stimulus choice. Stimuli were selected with participant input, considering their preferences and levels of success in naming, and lexical factors, such as word frequency, were held constant across stimulus conditions. It is possible, however, that category differences led to subtle differences in stimulus familiarity or concreteness that influenced participants differently or that certain categories are, for some reason, more amenable to this type of stimulation than others. In addition, typicality was not controlled in this study between categories or conditions. Given the role that typicality has been shown to play in within-category generalization of naming ability (Kiran, 2008; Kiran, Sandberg, & Sebastian, 2011; Kiran & Thompson, 2003), this may have influenced the outcomes of this study. Further research will be necessary to explore the interactions between category, word frequency, familiarity, typicality, and other lexical factors and response to masked priming. These issues may become clearer as additional data are collected in the course of this research project.

Last, considering stimulus factors as they might relate to the cross-category generalization observed, no particular patterns are evident in the current data set to suggest any strong relationship between particular categories and generalization across category boundaries; the two participants who showed cross-category generalization effects, however, were the two who also showed the strongest response to treatment (P1 and P3). This suggests that masked primes that are effective in obtaining a training effect may be effective in creating cross-category generalization as well regardless of the semantic categories involved. It should be noted, though, that one of these two participants (P3) was trained on two different subcategories of famous faces and showed the largest cross-category generalization effects. This may have been because the two categories were related to each other at a broad level (all famous people). If this were the case, however, then

other forms of within-category generalization (i.e., to UU items) should also have been noted. Given that this was not the case, it seems unlikely that the cross-category generalization effects noted were actually due to the items in the two categories being related.

Conclusion

The preliminary data presented here continue to support the idea that masked repetition priming can have a beneficial effect on naming in individuals with anomia due to aphasia. Differences between participants have provided early indicators of individual, protocol, and stimulus factors that must be considered and understood as this approach is investigated further. Additional studies are needed to refine the protocol, determine its best application, and compare its effectiveness and clinical feasibility with other treatment approaches and techniques known to improve naming in individuals with anomia. In addition, given that the results presented here reflect small functional changes, despite significant effect sizes, studies are needed to identify the best ways to apply this form of treatment to maximize both clinical and functional outcomes.

Acknowledgments

This work was supported by National Institute on Deafness and Other Communication Disorders Grant 5 R03 DC012643-02 (awarded to JoAnn P. Silkes). Thanks to Sara Pack, Amanda Hendricks, and Julie Cooke for assistance with stimulus development and data processing and the University of Washington Aphasia Research Lab for ongoing support.

References

- Andreewsky, E., & Seron, X. (1975). Implicit processing of grammatical rules in a classical case of agrammatism. *Cortex*, 11, 379–390.
- Avila, C., Lambon Ralph, M. A., Parcet, M., Geffner, D., & Gonzalez-Darder, J. (2001). Implicit word cues facilitate impaired naming performance: Evidence from a case of anomia. *Brain and Language*, 79, 185–200. doi:10.1006/brln.2001.2472
- Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology Review*, 16, 161–169. doi:10.1007/s11065-006-9013-7
- Boyle, M. (2004). Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. *American Journal of Speech-Language Pathology*, 13, 236–249.
- Boyle, M., & Coelho, C. A. (1995). Application of semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech-Language Pathology*, 4, 94–98.
- Brysaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41, 977–990. doi:10.3758/BRM.41.4.977
- Busk, P. L., & Serlin, R. (1992). Meta-analysis for single case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 187–212). Hillsdale, NJ: Erlbaum.
- Davies, M. (2008). *The Corpus of Contemporary American English: 450 million words, 1990-present*. Retrieved from <http://corpus.byu.edu/coca/>
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321. doi:10.1037/0033-295X.93.3.283
- Fillingham, J. K., Sage, K., & Lambon Ralph, M. A. (2005a). Further explorations and an overview of errorless and errorful therapy for aphasic word-finding difficulties: The number of naming attempts during therapy affects outcome. *Aphasiology*, 19, 597–614.
- Fillingham, J. K., Sage, K., & Lambon Ralph, M. A. (2005b). Treatment of anomia using errorless versus errorful learning: Are frontal executive skills and feedback important? *International Journal of Language & Communication Disorders*, 40, 505–523.
- Forster, K. I., Mohan, K., & Hector, J. (2003). The mechanics of masked priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked priming: The state of the art* (pp. 3–37). New York, NY: Psychology Press.
- Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language*, 43, 182–216.
- Francis, W. S. (2014). Repetition priming in picture naming: Sustained learning through the speeding of multiple processes. *Psychonomic Bulletin & Review*, 21, 1301–1308. doi:10.3758/s13423-014-0610-9
- Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York, NY: Wiley.
- Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *The Boston Naming Test*. Philadelphia, PA: Lea and Febinger.
- Kendall, D. L., Rosenbek, J. C., Heilman, K. M., Conway, T., Klenberg, K., Gonzalez Rothi, L. J., & Nadeau, S. E. (2008). Phoneme-based rehabilitation of anomia in aphasia. *Brain and Language*, 105, 1–17.
- Kertesz, A. (1982). *The Western Aphasia Battery*. New York, NY: Grune and Stratton.
- Kiefer, M. (2002). The N400 is modulated by unconsciously perceived masked words: Further evidence for an automatic spreading activation account of N400 priming effects. *Cognitive Brain Research*, 13, 27–39. doi:S0926641001000854 [pii]
- Kiran, S. (2008). Typicality of inanimate category exemplars in aphasia treatment: Further evidence for semantic complexity. *Journal of Speech, Language, and Hearing Research*, 51, 1550–1568. doi: 10.1044/1092-4388(2008/07-0038)
- Kiran, S., Sandberg, C., & Sebastian, R. (2011). Treatment of category generation and retrieval in aphasia: Effect of typicality of category items. *Journal of Speech, Language, and Hearing Research*, 54, 1101–1117. doi:10.1044/1092-4388(2010/10-0117)
- Kiran, S., & Thompson, C. K. (2003). The role of semantic complexity in treatment of naming deficits: Training semantic categories in fluent aphasia by controlling exemplar typicality. *Journal of Speech, Language, and Hearing Research*, 46, 773–787.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- LaPointe, L. L., & Horner, J. (1979). *Reading Comprehension Battery for Aphasia*. Tegoid, OR: G. C. Publications.
- Linebaugh, C. W., Shisler, R. J., & Lehner, L. (2005). CAC classics cueing hierarchies and word retrieval: A therapy program. *Aphasiology*, 19, 77–92.
- Martin, N., & Dell, G. S. (2007). Common mechanisms underlying perseverative and non-perseverative sound and word

- substitutions. *Aphasiology*, 21, 1002–1017. doi:10.1080/02687030701198346
- Mimura, M., Goodglass, H., & Milberg, W. (1996). Preserved semantic priming effect in alexia. *Brain and Language*, 54, 434–446.
- Nadeau, S. E. (2001). Phonology: A review and proposals from a connectionist perspective. *Brain and Language*, 79, 511–579.
- Nicholas, L. E., & Brookshire, R. E. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research*, 36, 338–350.
- Off, C. A., & Griffin, J. (2014, May). Treatment intensity and the effect of repetition priming on naming performance in individuals with aphasia. Poster presented at the 44th annual Clinical Aphasiology Conference, St. Simons Island, GA.
- Raven, J. C. (1976). *Coloured Progressive Matrices: Set A, Ab, B*. Oxford, UK: Oxford Psychologists Press.
- Regard, M., Strauss, E., & Knapp, P. (1982). Children's production on verbal and non-verbal fluency tasks. *Perceptual and Motor Skills*, 55, 839–844.
- Reingold, E. M. (2004). Unconscious perception: Assumptions and interpretive difficulties. *Consciousness and Cognition*, 13, 117–122.
- Revonsuo, A. (1995). Words interact with colors in a globally aphasic patient: Evidence from a stroop-like task. *Cortex*, 31, 377–386.
- Revonsuo, A., & Laine, M. (1996). Semantic processing without conscious understanding in a global aphasic: Evidence from auditory event-related brain potentials. *Cortex*, 32, 29–48.
- Roberts, D. J., Lambon Ralph, M. A., & Woollams, A. M. (2010). When does less yield more? The impact of severity upon implicit recognition in pure alexia. *Neuropsychologia*, 48, 2437–2446. doi:10.1016/j.neuropsychologia.2010.04.002
- Robey, R. R. (2004). A five-phase model for clinical-outcome research. *Journal of Communication Disorders*, 37, 401–411.
- Schmidt, R. A., & Lee, T. D. (2005). *Motor control and learning: A behavioral emphasis*. Champaign, IL: Human Kinetics.
- Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language*, 54, 199–227. doi:10.1016/j.jml.2005.10.002
- Schwartz, M. F., Saffran, E. M., Bloch, D. E., & Dell, G. S. (1994). Disordered speech production in aphasic and normal speakers. *Brain and Language*, 47, 52–88. doi:10.1006/brln.1994.1042
- Silkes, J. P., Brookshire, C. E., Dierkes, K., & Kendall, D. (2012, May). Masked priming treatment for anomia. Poster presented at the Clinical Aphasiology Conference, Lake Tahoe, CA.
- Silkes, J. P., Dierkes, K., & Kendall, D. (2012). Masked repetition priming effects on naming in aphasia: A Phase I treatment study. *Aphasiology*, 27, 381–397. doi:10.1080/02687038.2012.745475
- Turner-Stokes, L., Kalmus, M., Hirani, D., & Clegg, F. (2005). The Depression Intensity Scale Circles (DISCs): A first evaluation of a simple assessment tool for depression in the context of brain injury. *Journal of Neurology, Neurosurgery, & Psychiatry*, 76, 1273–1278. doi:10.1136/jnnp.2004.050096
- Tyler, L. K. (1992). The distinction between implicit and explicit language function: Evidence from aphasia. In A. D. Milner & M. D. Rugg (Eds.), *The neuropsychology of consciousness* (pp. 159–179). San Diego, CA: Academic Press.
- Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Research Methods, Instruments, & Computers*, 36, 481–487.

Appendix

Stimulus Lists

Table A1. Stimuli selected for P1.

Category	T	UE	UU
L1 Produce	avocado beets fig garlic honeydew olives papaya parsley rhubarb squash	artichoke asparagus banana chickpeas grapefruit kale leeks mango scallions zucchini	cauliflower coconut cucumber dates guava kiwi okra plums pumpkin yam
L2 Sports & games	air hockey Battleship cross-country ^a hunting juggling long jump Operation rugby skateboarding tetherball	archery bullfighting Chutes and Ladders Frisbee kayaking Ping-Pong rodeo shot put skiing Twister	canoeing dominoes gymnastics lacrosse Monopoly roller derby snowboarding sudoku surfing yoga

Note. T = trained; UE = untrained–exposed; UU = untrained–unexposed; L1 = list 1 (first category trained); L2 = list 2 (second category trained).

^aCross-country refers to running.

Table A2. Stimuli selected for P2.

Lists	T	UE	UU
L1 Occupations	butcher chef chemist coach farmer mechanic miner ranger surgeon veterinarian	artist diver pediatrician referee soldier tailor	conductor nurse painter reporter roofer umpire
L2 Vehicles	ambulance blimp canoe ferry gondola motorcycle raft wagon toboggan helicopter	cruise ship paddleboat sailboat school bus scooter snowmobile tow truck train trolley unicycle	golf cart van jet police car rocket

Note. T = trained; UE = untrained–exposed; UU = untrained–unexposed; L1 = list 1 (first category trained); L2 = list 2 (second category trained).

Table A3. Stimuli selected for P3.

Lists	T	UE	UU
L1 Famous faces (entertainment)	Alan Alda Angela Lansbury Carol Burnett Danny Kaye Dick Clark Eddie Murphy Jodie Foster Luciano Pavarotti Meg Ryan Ron Howard	Barbara Streisand Bette Midler Demi Moore Gene Hackman Judy Garland Kenny Rogers Marie Osmond Mel Gibson Michael Landon Sharon Stone	Anthony Hopkins Brooke Shields James Dean John Travolta Lily Tomlin Regis Philbin Roseanne Barr Rosie O'Donnell Susan Sarandon Whoopi Goldberg
L2 Famous faces (other)	Colin Powell Dan Rather Dennis Rodman Mother Theresa Nancy Reagan Peter Jennings Saddam Hussein Winston Churchill	Barbara Bush Dwight Eisenhower Henry Kissinger Madeline Albright Michael Jordan Nelson Mandela Newt Gingrich Ted Koppel	Bob Dole Charles Manson Janet Reno Jesse Ventura Mary Lou Retton Mikhail Gorbachev Ross Perot Tom Brokaw

Note. T = trained; UE = untrained–exposed; UU = untrained–unexposed; L1 = list 1 (first category trained); L2 = list 2 (second category trained).

Table A4. Stimuli selected for P4.

Lists	T	UE	UU
L1 Things to wear	apron clog goggles headband jersey mitten necklace pajamas shirt	blouse contacts pin scarf shorts sock suspenders sweatshirt vest	bracelet earmuffs earring helmet kilt mascara pants sandal shawl
L2 Sports & games	archery dominoes fencing horseshoes juggling pool rafting Scrabble tetherball Twister	bullfighting Frisbee hockey hunting jump rope long jump poker roller derby shot put volleyball	diving gymnastics javelin kickball lacrosse Monopoly rodeo rugby soccer surfing

Note. T = trained; UE = untrained–exposed; UU = untrained–unexposed; L1 = list 1 (first category trained); L2 = list 2 (second category trained).