

# When does word meaning affect immediate serial recall in semantic dementia?

ELIZABETH JEFFERIES

*University of Bristol, Bristol, England*

ROY JONES

*Research Institute for the Care of the Elderly, St. Martin's Hospital, Bath, England*

DAVID BATEMAN

*Royal United Hospital, Bath, England*

and

MATTHEW A. LAMBON RALPH

*University of Manchester, Manchester, England*

Patients with semantic dementia can show superior immediate recall of words that they still understand relatively well, as compared with more semantically degraded words, suggesting that conceptual knowledge makes a major contribution to phonological short-term memory. However, a number of studies have failed to show such a recall difference, challenging this view. We examined the effect of several methodological factors on the recall of known and degraded words in 4 patients with semantic dementia, in order to investigate possible reasons for this discrepancy. In general, our patients did exhibit poorer recall of the degraded words and made more phonological errors on these items. In addition, set size affected the magnitude of the recall advantage for known words. This finding suggests that semantic degradation influenced the rate of learning in the immediate recall task when the same items were presented repeatedly. The methods used to select known and degraded items also impacted on the recall difference. List length, however, did not affect the advantage for known words. The coherence of items in phonological short-term memory was affected by their semantic status, but not by the length of the material to be retained. The implications of these findings for the role of semantic and phonological representations in verbal short-term memory are discussed.

There is considerable debate about the extent to which phonology and semantics are functionally independent. The autonomy of these systems has been called into question by Patterson, Graham, and Hodges (1994), who argued that the semantic system provides an important source of constraint on phonological activation, particularly in demanding tasks such as immediate serial recall (ISR). According to this theory, every time a word is spoken or comprehended, semantic activation necessarily co-occurs with the activation representing the phoneme sequence for that word and, as a result, increases the likelihood that the phonological elements of the word will be produced in the correct configuration in ISR. Other authors have maintained that the integrity of phonological representations is not dependent on input from the semantic system (McCarthy & Warrington, 2001).

Patterson et al.'s (1994) contention is consistent with research suggesting that ISR tasks draw on the full range

of representations involved in language processing. The importance of phonological codes in ISR has long been recognized (Baddeley & Hitch, 1974; Conrad, 1964; Conrad & Hull, 1964), but lexical and semantic representations also appear to make a major contribution to the recall of familiar words. Normal participants recall words better than nonwords (Hulme, Maughan, & Brown, 1991; Hulme, Roodenrys, Brown, & Mercer, 1995) and recall frequently encountered words better than low-frequency words (Hulme et al., 1997; Watkins & Watkins, 1977). In addition, concrete/highly imageable words are recalled better than abstract/low-imageability words (Bourassa & Besner, 1994; Walker & Hulme, 1999), and recall improves when items are grouped by semantic category (Poirier & Saint Aubin, 1995).

Neuropsychological evidence concurs with the notion that an interaction between phonology and semantics underpins verbal short-term memory (STM). R. C. Martin and colleagues have described several patients with a reduced advantage for recalling words over nonwords, who purportedly had a specific difficulty in the retention of lexical-semantic information (R. C. Martin & Lesch, 1996; R. C. Martin, Lesch, & Bartha, 1999; R. C. Mar-

---

The first author was supported by an ESRC studentship. Correspondence concerning this article should be addressed to E. Jefferies, Department of Psychology, University of Manchester, Oxford Road, Manchester, M13 9PL, England (e-mail: beth.jefferies@man.ac.uk).

tin, Shelton, & Yaffee, 1994). In addition, N. Martin and Saffran (1997) reported that aphasic patients' semantic abilities, but not their phonological skills, were related to the magnitude of frequency and imageability effects in ISR, suggesting that semantic representations play a role in verbal STM, in addition to a phonological code. In line with this, ISR is impaired in patients with semantic dementia (SD), who show a specific decline in semantic memory. SD is the temporal variant of frontal-temporal dementia and is associated with progressive focal atrophy of the anterior inferolateral regions of the temporal lobe (Snowden, Goulding, & Neary, 1989). SD patients are anomia and have impaired comprehension on both verbal and nonverbal tasks. However, their perceptual and spatial skills, new episodic learning, nonverbal reasoning, syntax, and phonology remain largely intact (Hodges, Patterson, Oxbury, & Funnell, 1992). SD patients almost never produce phonological errors in spontaneous speech, have intact digit span, and generally perform well on phonological tasks such as minimal pair discrimination (Knott, Patterson, & Hodges, 1997). In contrast, all SD patients show a pattern of phonological breakdown in ISR, in which phonemes migrate to new positions in the list (McCarthy & Warrington, 1987; Patterson et al., 1994). For example, Patterson et al. reported that the onsets in the words *mint* and *rug* were exchanged in ISR to produce the responses *rint* and *mug*. Normal participants show a similar pattern of phonological migration errors in their serial recall of nonwords, which, by definition, lack lexical and semantic representations (Treiman & Danis, 1988).

If stable semantic representations play a role in maintaining the phonological coherence of words in STM, the likelihood of phonological disintegration for a given word should be associated with the degree to which it is semantically degraded. A number of studies have selected, from such semantic tests as naming and word-picture

matching, words that individual patients understand relatively well or poorly, in order to test this prediction. Patterson et al. (1994) found very substantial recall differences between well-known and more semantically degraded words for 3 patients with SD. Although the known and the degraded words used in this study were not matched for frequency, significant differences remained when the known items with the highest frequency were discarded. Several subsequent studies have shown a substantial recall difference between known and degraded words matched for frequency on an item-by-item basis (Jefferies, Jones, Bateman, & Lambon Ralph, in press; Knott et al., 1997; Knott, Patterson, & Hodges, 2000). In addition to these studies of SD patients, ISR differences between relatively known and degraded words have been reported in semantically impaired patients following cardiovascular accidents (Forde & Humphreys, 2002) and herpes simplex encephalitis (Caza, Belleville, & Gilbert, 2002).

Although ISR differences between known and degraded words do occur, there have been some notable failures to find such differences (Funnell, 1996; Lambon Ralph & Howard, 2000; McCarthy & Warrington, 1987, 2001; Warrington, 1975), and their interpretation remains controversial (see Table 1). Knott et al. (1997) did not find a significant recall difference between known and degraded words in one patient (B.M.), despite finding a difference in a second patient (A.M.). In addition, McCarthy and Warrington's (2001) patient M.N.A. was able to recall a normal number of words that she did not understand. These findings appear to challenge Patterson et al.'s (1994) assertion that semantics plays a major role in maintaining the phonological coherence of words in STM. As an alternative, McCarthy and Warrington (2001) argued that verbal STM could operate without the involvement of semantics and that additional phonological-lexical impairments were responsible for the known-degraded recall differences observed in some studies. It

**Table 1**  
Previous Studies That Have Examined Immediate Serial Recall of Known and Degraded Words in Patients With Semantic Dementia

Study	Patient	Set Size	Frequency Matched	Mean Known Items Recalled (%)	Mean Degraded Items Recalled (%)	Known-Degraded Difference ( <i>p</i> )
Patterson, Graham, & Hodges (1994)	J.L.	60	no	83	50	< .001
Patterson et al. (1994)	P.P.	36	no	85*	57*	< .001
Patterson et al. (1994)	F.M.	36	no	97	74	< .001
McCarthy & Warrington (2001)	M.N.A.	30	yes	66	69	n.s.
Knott, Patterson, & Hodges (1997)	A.M.	24	yes	85	68	< .01
Knott et al. (1997)	B.M.	24	yes	66	58	.09
Knott, Patterson, & Hodges (2000)	F.M.	20	yes	67	48	< .001
Warrington (1975)	A.B.	15	no	88	80	n.s.
Warrington (1975)	E.M.	15	no	92	84	n.s.
McCarthy & Warrington (1987)	N.H.B.	12	no	90	90	n.s.
Lambon Ralph & Howard (2000)	I.W.	10	yes	span = 2.98 <sup>†</sup>	span = 2.82 <sup>†</sup>	n.s.
Funnell (1996)	E.P.	7	yes	69	86	n.s.

Note—The studies are arranged according to set size. When data from more than one list length or testing session were available, the average was calculated. \*Includes scores from two testing sessions in which the classification of words as known and degraded was up to date. <sup>†</sup>Tested using a staircase method in which list length depended on previous recall accuracy.

is important to note, however, that the patients who failed to show a difference in recall accuracy between known and degraded words still made an abnormal number of phonological errors in ISR. Moreover, the inconsistency in the size of the recall accuracy difference between known and degraded words could be a consequence of discrepancies in methodology. Consequently, in the research presented here, we investigated the effect of various methodological factors on the size of the known–degraded recall difference in SD patients.

We focused on four methodological variables that could influence the size of the known–degraded recall difference: (1) the method used to classify items as known and degraded, (2) the length of the lists to be recalled, (3) the frequency matching of known and degraded words, and (4) the total number of known and degraded words in the lists (set size). Each of these factors will be discussed in turn.

### Method Used to Obtain Known and Degraded Words

In previous studies, a wide variety of methods have been used to select known and degraded items for recall, including picture naming, word–picture matching, definitions, verbal fluency, and spontaneous speech. Since the semantic degradation underlying the known–degraded distinction varies continuously, the point of cutoff between *known* and *degraded* items may differ across these methods. When particularly difficult semantic tests are used, items that are selected as *degraded* may still be understood to a certain extent, whereas when easier tests are used, items that are selected as *known* may have lost the finer nuances of their meaning. Consequently, the method that is adopted to select the known and the degraded words may affect the size of the comprehension difference between them. The choice of method may be especially critical for patients who are particularly impaired at certain types of tests. For example, Knott et al.'s (1997) patient B.M. was particularly poor at pictorial tasks, and the tasks used to select his known and degraded words were pictorial in nature. Consequently, his *degraded* words may have been relatively well known, at least as assessed by verbal tests. In the present study, two different methods for obtaining known and degraded items were compared, providing some measure of the sensitivity of the known–degraded recall difference to this variable.

### List Length

SD patients almost never produce phonological errors in spontaneous speech and only occasionally produce them in single-word repetition, perhaps because the STM system is not sufficiently taxed by these tasks. Consequently, longer lists may increase the likelihood of phonological breakdown for degraded words. On very long lists, however, phonological errors may fall away if few of the correct phonemes are maintained until output. Knott et al. (1997) partly attributed the small number of phono-

logical errors made by B.M. to list length. He was tested on six-word lists and largely made omission errors. Previous studies have not manipulated list length systematically and have generally tested different patients at a single list length, set according to their word span. In this study, list length was manipulated for each patient, allowing an investigation of this factor on the occurrence of phonological errors for known and degraded words.

### Frequency Matching

Patterson et al. (1994) obtained large known–degraded recall differences for 3 patients, but they did not match the items for word frequency. No control data were reported, making it difficult to gauge how much of the recall difference was due to the known and degraded status of the words and how much corresponded to the standard frequency effect observed in normal performance (e.g., Hulme et al., 1997). Although it is clearly problematic not to match for frequency, this process has its own inherent difficulties. First, frequency matching may use up some of the natural variation in the known–degraded dimension, since the two factors are correlated and lower frequency items generally degrade earlier in the course of the disease (Funnell, 1995). By matching closely for frequency, therefore, one is unable to maximize the known versus degraded difference. Second, since known words typically have higher frequencies than do degraded words, frequency matching will be possible only for a small proportion of items, leading to small set sizes. Third, frequency matching can largely be achieved for medium-frequency words only. Finally, there are likely to be personal oddities in word frequency, since words that individuals use regularly because of their occupations or interests will have higher personal frequencies than those obtained from the database counts. Therefore, *frequency-matched* pairs may not be genuinely matched and, instead, may be governed by these personal oddities. In this study, the outcomes of experiments that did and did not match for frequency were compared.

### Set Size

Funnell (1996) failed to find an ISR difference between known and degraded words when there were seven words in each category and suggested that the small size of the word pool might have accounted for this null result. Some support for this suggestion is provided in Table 1, which lists previous studies in which the recall of known and degraded words in SD patients, arranged according to set size, has been examined. It is clear that studies involving larger sets of known and degraded words obtained significant known–degraded recall differences more often than did those involving smaller set sizes (four out of six experiments vs. one out of six). There is, in fact, a significant correlation of .73 between set size and study outcome ( $p = .007$ ). In experiments with small set sizes, the same items are presented repeatedly, making them easier to identify, retain, and produce at recall. In line with this suggestion, ISR is higher

in normal participants when the items on each trial are drawn from a small pool and presented repeatedly (V. Coltheart, 1993; Conrad, 1963). Set size may affect the recall of degraded words to a greater extent than known words because patients can become more familiar with the phonological forms of degraded words as they are repeated, allowing them to catch up with the recall of known words. Roodenrys and Quinlan (2000) found larger frequency effects for healthy participants when they were tested in an open-set condition in which items were never repeated, suggesting that lexical and semantic factors might play a diminished role in verbal STM when set size is small. Contrary to this suggestion, however, Knott et al. (1997) found that although set size affected recall accuracy in their patient A.B., the effect of imageability on ISR did not differ for small and large set sizes. In the work presented here, the effect of set size on the recall of known and degraded words was examined.

### CASE DESCRIPTIONS

In this work, we examined 4 SD patients, who are described in order of severity. A summary of the background neuropsychological assessment is shown in Table 2. This research conformed to the ethical guidelines of the American Psychological Association. E.K. was a 60-year-old right-handed woman who had left school at the age of 15 and had been experiencing worsening word-finding difficulties for around 5 years. She was living alone and doing occasional cooking and cleaning jobs at the time of the study. An MRI scan showed bilateral temporal lobe atrophy that was more marked in the left hemisphere. Her neuropsychological profile was dominated by a mod-

erate impairment of semantic memory. She performed poorly on tests requiring comprehension of words and pictures—for example, word–picture matching and the Pyramids and Palm Trees Test (Howard & Patterson, 1992). She was severely anomia in spontaneous speech, word fluency tasks, and confrontational picture naming. Her naming errors were predominantly omissions and semantic paraphasias. In common with other SD patients, she produced surface dyslexic errors in reading aloud and surface dysgraphic errors in spelling tasks. In contrast to her marked semantic difficulties, she was well oriented in time and place, had excellent episodic memory for recent events, and had no difficulty in remembering appointments. She performed normally on tests of visuospatial processing from the Visual Object and Space Perception Battery (VOSP; Warrington & James, 1991), and she was able to produce a good immediate copy of the Rey complex figure (Lezak, 1976). Her nonverbal reasoning on the Colored Progressive Matrices Test (Raven, 1962) was normal. Her speech was fluent and syntactically well formed, despite her anomia. She had intact single-word phonology, and she did not make phonological errors in her spontaneous speech or picture naming. She had normal spatial STM, as assessed by the Corsi block-tapping task, and normal verbal STM, as measured by forward and backward digit span (Wechsler, 1987). Her word span performance, however, was characterized by frequent phonological errors similar to those described in Patterson et al. (1994).

G.T., a 71-year-old right-handed male, left school at 14 years of age and worked as a builder and a technician in a higher education college. At the time of the study, he had been experiencing a gradual decline in his word

**Table 2**  
**Background Neuropsychological Scores**

Test	Maximum	E.K.	G.T.	P.D.	M.K.	Controls	
						<i>M</i>	<i>SD</i>
MMSE <sup>1</sup>	30	27	26	13*	21*	Cutoff > 24	–
Raven's matrices <sup>2</sup>	36	33	35	25*	22*	–	–
Digit span: forward <sup>3</sup>	–	6	6	7	5	6.8	0.9 <sup>a</sup>
Digit span: backward <sup>3</sup>	–	7	4	5	4	4.7	1.2 <sup>a</sup>
Spatial span: forward <sup>4</sup>	–	6	5	–	5	5–6 <sup>b</sup>	–
Naming	64	17*	11*	4*	2*	62.3	1.6 <sup>a</sup>
Word–picture matching	64	46*	32*	17*	11*	63.7	0.5 <sup>a</sup>
PPT: pictures <sup>5</sup>	52	35*	37*	26*	33*	51.1	1.1 <sup>a</sup>
PPT: words <sup>5</sup>	52	36*	32*	26*	26*	51.2	1.4 <sup>a</sup>
Category fluency: living	–	13*	13*	5*	0*	60.3	12.6 <sup>a</sup>
Category fluency: man-made	–	18*	11*	2*	1*	54.8	10.3 <sup>a</sup>
Letter fluency (F, A, S)	–	29	24	22	2*	44.2	11.2 <sup>a</sup>
Rey figure immediate copy <sup>6</sup>	36	34	34	36	30	34.0	3.0 <sup>c</sup>
VOSP: incomplete letters <sup>7</sup>	20	20	18	3*	10*	19.2	0.8 <sup>a</sup>
VOSP: dot counting <sup>7</sup>	10	10	10	10	10	9.9	0.3 <sup>a</sup>
VOSP: position discrimination <sup>7</sup>	20	20	20	16*	17*	19.8	0.6 <sup>a</sup>
VOSP: cube analysis <sup>7</sup>	10	10	10	5*	6	9.7	2.5 <sup>a</sup>

\*Denotes abnormal performance. <sup>1</sup>MMSE, Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975). <sup>2</sup>Raven's Colored Progressive Matrices (Raven, 1962). <sup>3</sup>Wechsler Memory Scale—Revised (Wechsler, 1987). <sup>4</sup>Wechsler Memory Scale—III (Wechsler, 1997). <sup>5</sup>PPT, Pyramids and Palm Trees Test (Howard & Patterson, 1992). <sup>6</sup>Taken from Lezak, 1976. <sup>7</sup>VOSP, Visual Object and Space Perception Battery (Warrington & James, 1991). <sup>a</sup>Control data from Bozeat, Lambon Ralph, Patterson, and Hodges (2002). <sup>b</sup>Normal range for age-matched participants. <sup>c</sup>Control data from N. L. Graham, Patterson, and Hodges (2001).

finding and comprehension for 5 years. An MRI scan showed marked bilateral circumscribed temporal lobe atrophy. His cognitive profile was similar to that of E.K., although his semantic impairments were a little more severe. He was impaired on a range of pictorial and verbal tests of semantic memory. In contrast, he was well oriented in time and space and had intact visuospatial skills, nonverbal reasoning abilities, and memory for recent events. His speech was fluent and syntactically well formed but was characterized by anomia and frequent circumlocutions, and his conversation was repetitive. He did not make phonological errors in spontaneous speech or picture naming. He had good verbal STM, as measured by forward and backward digit span, although his word span performance was characterized by frequent phonological errors. His hearing was slightly impaired in his right ear.

P.D., a 73-year-old right-handed woman, left school at the age of 14 and later worked as a regional organizer for a large charity. She had an 8-year history of worsening semantic memory problems, and these were very severe at the time of testing. An MRI scan showed very marked bilateral temporal lobe atrophy that was worse in the right hemisphere, with relative preservation of more medial temporal lobe structures including the hippocampus, and also evidence of some more generalized cortical atrophy. P.D. was near floor on a range of tests that required comprehension of pictures and words. Early in the course of the disease, she had experienced particular problems with recognizing objects and people, and at the time of testing, she showed poorer performance on pictorial, as compared with verbal semantic, tests, consistent with her predominantly right-sided atrophy (Evans, Heggs, Antoun, & Hodges, 1995). Although she had been well oriented for time and place when she first presented in 1996, she was more poorly oriented at the time of testing and occasionally became lost. She also showed some impairment in visuospatial skills and nonverbal reasoning. P.D. exhibited some behavioral changes, including disinhibition, which would be consistent with the disease process affecting basal frontal as well as temporal regions (Snowden, Neary, & Mann, 1996). She became increasingly difficult to test and withdrew from the study before all the experiments reported here had been completed.

M.K., a 67-year-old right-handed woman, was the most severely semantically impaired patient included in the study. She left school at the age of 17 and had previously been employed in clerical work. Her family reported a 3-year history of worsening semantic problems. An MRI scan showed marked temporal lobe atrophy that was strongly lateralized to the left side. She performed at or near floor on tests of semantic memory. In contrast to her semantic impairments, she remained well oriented in time and place, and her memory for recent events was excellent. Her verbal STM was normal, as assessed by forward and backward digit span. At the time of testing, she appeared to have good single-word phonology and did not produce phonological errors in spontaneous

speech or picture naming. She was impaired on tests of nonverbal reasoning and visuospatial processing, but she did not show signs of disinhibition or other behavioral changes.

## EXPERIMENT 1

### Immediate Serial Recall of Frequency-Matched Known and Degraded Words Defined by Naming and Definitions

If semantic representations make an important contribution to verbal STM, ISR should be better for words that are still relatively well understood than for words whose meanings have become degraded. To test this prediction, known and degraded words were selected for each patient using two methods: picture naming and definitions in Experiment 1 and synonym judgment in Experiment 2. The patients were tested on a variety of list lengths in both experiments, in order to investigate the effect of load on the phonological coherence of degraded words.

#### Method

The patients were asked to name 80 pictures from the Snodgrass set, as well as 13 colors and 11 body parts, and to provide definitions for the same items. Attempts at naming were considered to be correct when the patients produced the right specific label for a picture. Definitions were considered to be correct when they contained enough specific information to allow the item to be identified from its description; gestures—for example, pointing at an item—were also accepted. For E.K., G.T., and P.D., items that were both named and defined correctly were classified as known, and items that were neither named nor defined correctly were classified as degraded. For M.K., this method did not produce a sufficient number of known words, and consequently, content words that she used correctly in her descriptions of eight complex pictures, including the *cookie theft* picture, were also included.

The known and the degraded words were matched for word frequency as closely as possible on an item-by-item basis, using data from Celex (Baayen, Piepenbrock, & van Rijn, 1993) and the MRC psycholinguistic database (M. Coltheart, 1981). It did not prove possible to match every pair of words for syllable length, but the average numbers of syllables in the known and the degraded words were as similar as possible. The constrained way in which the words were judged to be known and degraded and the item-by-item frequency matching of known to degraded words resulted in small set sizes for all 4 patients. Table A1 in the Appendix gives set size, along with mean word frequency, length, and imageability ratings, for each patient's known and degraded words.

Lists of known and degraded words were assembled by selecting items at random without replacement until all the items had been used and then repeating this process as required. The frequency-matched known and degraded word pairs were yoked so that they appeared in the same positions within corresponding lists. The patients were tested on lists containing three, four, five, and six words, although P.D. was not tested on three-word lists, due to time constraints, and M.K. was tested on two, three, four, and five words, because her performance was poorer than that of the other patients. There were 10 lists of known and degraded words at each length, and they were presented in a blocked fashion, using an ABBA design to control for practice effects. E.K. and G.T. were tested twice on lists containing four, five, and six words, and M.K. was tested twice on lists containing four and five words, in order to increase

the amount of data available for analysis. The repeated lists were separated from the original testing by a period of several weeks. Three healthy control participants were matched to each patient on the basis of sex, age, and years of education. Each patient's controls were tested on the same lists as the patient and also on lists up to seven words long, constructed in the same way. As the known and degraded items were specific to each patient, the different patients and their controls were tested on different sets of items. In this and subsequent experiments, items were read aloud at a rate of one word per second for immediate serial recall.

## Results

**Recall accuracy.** Both list and item recall can provide a measure of recall accuracy. The two methods generally produce the same pattern of results, although floor and ceiling effects can raise fewer problems for item recall. Therefore, in the interests of brevity, only item recall will be reported for every experiment.

Table 3 shows the number of items recalled in the correct order at each list length. A series of *t* tests was used to determine whether the patients performed significantly more poorly than their lowest scoring controls, combining across the different list lengths at which both the patients and the controls were tested. E.K.'s performance on the known words was not impaired [ $t(29) < 1$ ]. In contrast, her recall of the degraded words was significantly poorer than that of her lowest scoring control [ $t(29) = 4.23, p < .0001$ ]. Similarly, G.T.'s recall of the known words was actually better than his lowest scoring control [ $t(29) = 2.65, p < .05$ ], but he was markedly impaired on the degraded words [ $t(29) = 3.23, p < .01$ ]. M.K. showed the largest ISR impairment, perhaps because her semantic deficits were particularly severe. Her recall was substantially impaired for both known words [ $t(29) = 7.08, p < .0001$ ] and degraded words [ $t(29) = 8.42, p < .0001$ ]. In contrast to the other patients, P.D.'s ISR performance was not impaired. In fact, her recall

was significantly better than that of her poorest performing control, for both known words [ $t(29) = 3.59, p < .001$ ] and degraded words [ $t(29) = 2.03, p = .05$ ].

Two of the 4 patients displayed a significant recall advantage for the known words over the degraded words, consistent with the notion of a semantic contribution to verbal STM. G.T. showed a very substantial recall difference between known and degraded words when the data were combined across list lengths [ $t(132) = 4.13, p < .0001$ ]. M.K. also recalled a larger number of known than of degraded words [ $t(112) = 2.56, p < .05$ ]. In contrast, no difference between the recall of known and degraded words was found for E.K. [ $t(134) = 1.42, n.s.$ ] or P.D. [ $t(54) < 1$ ].

None of the control participants showed superior recall of the known words, suggesting that the results obtained for G.T. and M.K. genuinely reflected the involvement of semantics in ISR and not differences in difficulty between the two sets of words. For the most part, the controls showed no difference between the known and the degraded words [ $t(76-98) < 1.56, n.s.$ ], although 2 control participants showed superior recall of the degraded words [ $t(75) = 2.48, p < .05$ , and  $t(78) = 2.15, p < .05$ ]. This unexpected result may reflect the fact that the known and the degraded words were not perfectly matched on every characteristic affecting verbal STM and, when it was not possible to find a match, the degraded items were generally selected to be easier.

Table 3 shows that the percentage of recall of the known and degraded words decreased as list length was increased. List length affected the patients and the control participants in a similar way and did not systematically affect the size of the known-degraded recall difference.

**Error analysis.** The errors made by the patients and the controls were classified as belonging to one of six categories. Omission errors occurred when fewer items

**Table 3**  
Recall of Frequency-Matched Known and Degraded Words  
Defined by Naming and Definitions (Experiment 1)

Patient	Words	List Length					
		2	3	4	5	6	7
E.K.	known	–	90.0	78.8*	64.0	55.0	–
	degraded	–	86.7	73.8*	59.0	50.0	–
E.K. controls (minimum)	known	–	–	80.0	60.0	51.7	48.6
	degraded	–	–	87.5	76.0	58.3	52.9
G.T.	known	–	93.3	90.0	70.0	60.8	–
	degraded	–	83.3	62.5*	60.0	42.5	–
G.T. controls (minimum)	known	–	–	87.5	64.0	41.7	44.3
	degraded	–	–	82.5	70.0	61.7	50.0
P.D.	known	–	–	90.0	88.0	70.0	–
	degraded	–	–	85.0	74.0	80.0	–
P.D. controls (minimum)	known	–	–	82.5	58.0	51.7	42.9
	degraded	–	–	87.5	68.0	56.7	61.4
M.K.	known	70.0	70.0	57.5*	50.0*	–	–
	degraded	65.0	60.0	43.0*	38.0*	–	–
M.K. controls (minimum)	known	–	100	85.0	70.0	63.3	57.1
	degraded	–	96.7	90.0	68.0	68.3	57.1

Note—Figures indicate the percentage of items recalled in the correct order. \*Recall below minimum score obtained across all control participants on both known and degraded words.

were recalled than were presented. Order errors occurred when an item was identical to one of the target items but was produced in the wrong place in the sequence. Repetition errors were target items recalled more than once. Intrusion errors were previously presented items recalled in the wrong list. Phonological errors contained at least half of the phonemes in a target word. Unrelated errors did not fit into any of the previous categories and were largely accounted for by patient responses that did not overlap sufficiently with the target words to reach the criterion for a phonological error.

Table 4 indicates the proportion of errors in each of these categories for known and degraded words, combining across list lengths. There were far more errors in the phonological and unrelated categories for the patients than for the controls. Chi-square analyses showed that different types of errors occurred on the known and the degraded words for G.T. [ $\chi^2(5) = 27.62, p < .0001$ ], E.K. [ $\chi^2(5) = 28.93, p < .0001$ ], and P.D. [ $\chi^2(5) = 13.41, p < .01$ ]. The standardized residuals were particularly high for phonological errors (range = 1.7–3.3), suggesting that this error category made a major contribution to the chi-square outcome. In contrast, M.K. did not make significantly different types of errors on the known and the degraded words [ $\chi^2(5) = 8.24, n.s.$ ], perhaps because a substantial number of phonological errors occurred in her recall of the known words, as well as of the degraded words.

In contrast with the patients, the control participants did not make a larger number of phonological errors in their recall of the degraded words and made very few phonological errors in either condition. Ten of the 12 control participants did not show any significant difference in errors between the known and the degraded words [ $\chi^2(5) = 1.85$ – $5.85$ , all *n.s.*]. Two control participants did show a significant difference [ $\chi^2(5) = 9.57, p < .05$ , and  $\chi^2(5) = 11.30, p < .05$ ], but this was apparently due to differences in the number of order and intrusion errors, rather than phonological errors.

The numbers of phonological and nonphonological errors were compared across short and long lists for E.K., G.T., and M.K. (see Figures 1A, 1B, and 1C). The *short* lists contained three and four items for E.K. and G.T. and two and three items for M.K. The *long* lists contained five

and six items for E.K. and G.T. and four and five items for M.K. P.D. was excluded from the analysis because there were insufficient errors on the shorter lengths to analyze. For all 3 patients, phonological errors appeared in large numbers on the shorter lists and did not increase substantially with list length. Percentage recall decreased as list length was increased, largely because the number of nonphonological errors (predominantly omissions) rose sharply. Consistent with this pattern, phonological errors accounted for a greater proportion of the total errors on short lists than on long lists in degraded word recall, for E.K. [ $\chi^2(1) = 15.24, p < .0001$ ], G.T. [ $\chi^2(1) = 7.90, p < .01$ ], and M.K. [ $\chi^2(1) = 6.75, p < .01$ ]. The difference in the proportion of phonological to nonphonological errors on short and long lists did not reach significance for known words [E.K. and M.K.,  $\chi^2(1) < 1$ ; G.T.,  $\chi^2(1) = 1.93, n.s.$ ], presumably because the number of phonological errors was much smaller.

## Discussion

Two patients, E.K. and G.T., showed impaired recall of degraded words, but not of known words, relative to control performance. G.T. showed the predicted recall difference between known and degraded words, but this did not reach significance for E.K. A 3rd patient, M.K., was markedly impaired at recalling both known and degraded words, although she nevertheless showed a significant recall difference between them. M.K. was the most semantically impaired patient in this study, and her comprehension of the known words may have been substantially impaired, although still superior to her comprehension of the degraded words. In contrast, P.D.'s recall accuracy was at a normal level for both known and degraded words, despite her severe semantic impairments. In this respect, she was similar to McCarthy and Warrington's (2001) patient, M.N.A., who was characterized as "repeating without semantics." An analysis of the errors made by the patients and controls, however, revealed that although P.D.'s accuracy remained at a normal level, like M.N.A.'s, the errors that she made were anything but normal. In our study, the number of phonological errors was much larger in the patients than in the controls, particularly for degraded words, suggesting that semantic impairment does affect the phonological coherence of items in STM,

**Table 4**  
Errors on Frequency-Matched Known and Degraded Words Defined by Naming and Definitions (Experiment 1)

Patient	Phonological		Unrelated		Omission		Order		Repetition		Intrusion	
	Known	Degraded	Known	Degraded	Known	Degraded	Known	Degraded	Known	Degraded	Known	Degraded
E.K.	.09*	.38*	.04*	.05*	.41	.28	.28	.19	.09	.06	.10	.05
G.T.	.27*	.59*	.10*	.10*	.42	.20	.09	.02	.07	.06	.05	.03
P.D.	.26*	.67*	.03*	.03	.19	.03	.16	.06	.32	.15	.03	.06
M.K.	.64*	.65*	.16*	.24*	.08	.06	.03	.00	.03	.03	.05	.02
All controls (maximum)	.04	.06	.02	.04	.81	.64	.43	.48	.29	.26	.37	.34

Note—The errors in each category are expressed as a proportion of the total number of errors across all list lengths. \*Patient scores were larger than the maximum observed for controls.

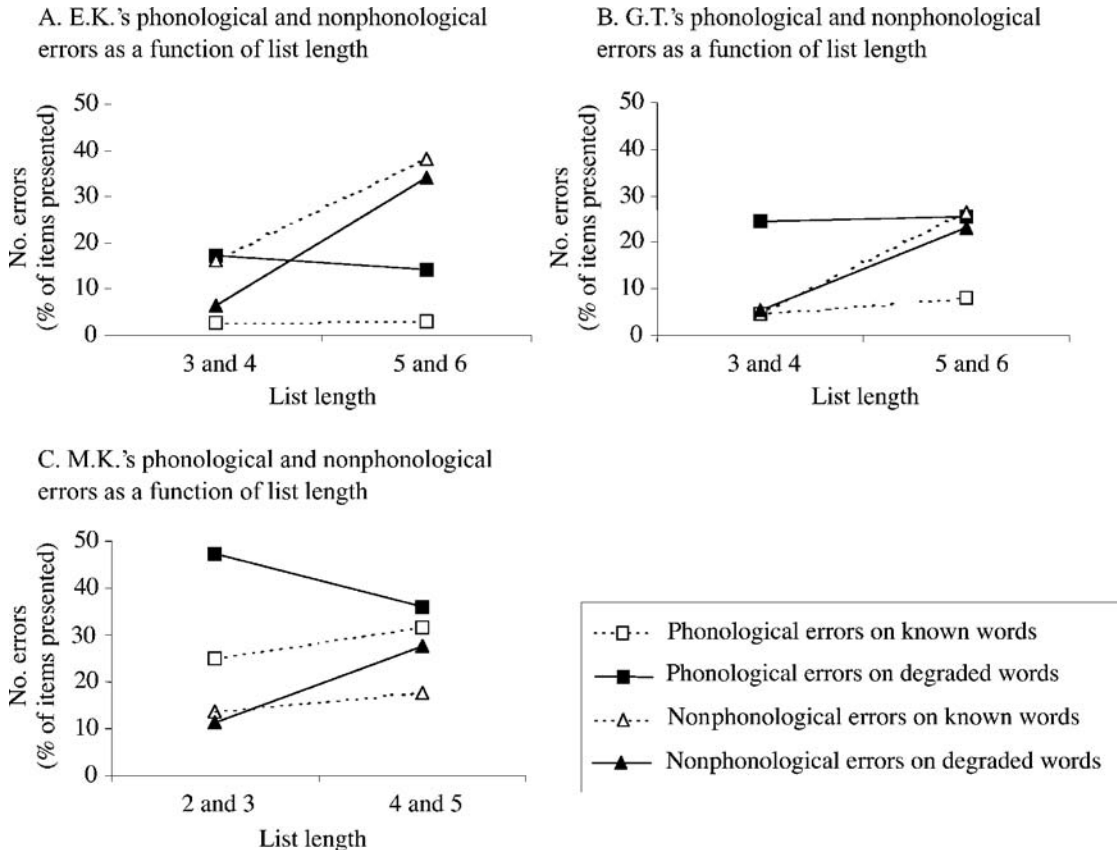


Figure 1. Phonological and nonphonological errors on known and degraded words defined by naming and definitions as a function of list length (Experiment 1).

as predicted by Patterson et al. (1994). Although the patients made more frequent phonological errors on the degraded words, they also made an abnormal number of these errors on the known words. It seems likely that these errors reflected the fact that a continuum of semantic impairment underlies the known–degraded distinction. Although the known words were selected to be better understood than the degraded words, the semantics of these words may not have been entirely intact, since the patients could have passed the naming and definitions tests without understanding the finer nuances of the words' meanings. Therefore, phonological errors on both sets of items may have resulted from semantic impairment.

Recall declined as list length was increased but declined at a similar rate for known and degraded words. The size of the known–degraded difference did not vary consistently with length, as long as recall was off floor and ceiling. However, list length did appear to affect the proportion of phonological to nonphonological errors. Nonphonological errors, predominantly omissions, increased with length for both known and degraded words, but phonological errors occurred frequently for degraded words, even on very short lists, and did not increase markedly with length.

## EXPERIMENT 2 Immediate Serial Recall for Frequency-Matched Known and Degraded Words Defined by Synonym Judgments

In this experiment, a second set of known and degraded words was selected using a synonym judgment task, in order to establish whether the results of the first experiment would replicate regardless of the change in the method used to select the items.

### Method

A synonym judgment task was used to produce lists of known and degraded words for E.K., G.T., and P.D. The participants were asked which of three words was closest in meaning to a target word (for example, "which word is closest in meaning to suffix: inflection, temerity, or perpetrator?"). The test was administered twice on two separate occasions. Known items were selected from the consistently correct trials, and degraded items were defined as those trials in which performance was consistently incorrect. M.K. was not included, since she performed too poorly on the synonym judgment task to produce enough known words.

The frequency of the known and the degraded words was matched on an item-by-item basis, and the groups were matched for word length, as has been described for Experiment 1. Again, the set sizes were small for all 4 patients. Table A2 in the Appendix gives set size,



along with mean frequency, length, and imageability for the known and the degraded words selected for each patient. Lists of known and degraded words were assembled in the same way as that in Experiment 1. G.T. and E.K. were tested on lists containing two, three, four, and five items. P.D. was tested on three- to five-item lists. The control participants were tested on three-, four-, five-, and seven-item lists. There were 10 lists of known and degraded words at each length, and they were presented in a blocked fashion, using an ABBA design.

## Results

**Recall accuracy.** Table 5 shows the percentages of items recalled in the correct order at each list length by the patients and the controls. In general, these words were recalled more poorly than the words in Experiment 1, perhaps because they were more abstract in nature.

A series of *t* tests was used to determine whether the patients had performed significantly more poorly than their lowest scoring controls, combining across the different list lengths at which both patients and controls were tested. G.T.'s recall of the known words was not impaired [ $t(29) = 1.55$ , n.s.], but his recall of the degraded words was substantially impaired [ $t(29) = 6.94$ ,  $p < .0001$ ]. E.K.'s recall was impaired for both known words [ $t(29) = 4.87$ ,  $p < .0001$ ] and degraded words [ $t(29) = 5.22$ ,  $p < .0001$ ]. In contrast, P.D.'s recall was not impaired for either known words or degraded words [ $t(29) < 1$ ], as in Experiment 1.

G.T. recalled the known words much more accurately than the degraded words [ $t(78) = 4.02$ ,  $p < .0001$ ], but there was no difference in recall between the known and the degraded words for E.K. [ $t(77) < 1$ ] and P.D. [ $t(58) < 1$ ]. Therefore, this experiment replicated the findings of Experiment 1: The same patients showed a significant known-degraded difference, even when the words were selected using a different method. None of the controls showed a significant difference between the known and the degraded words [all  $t_s(77-98) < 1.39$ , n.s.].

**Error analysis.** Errors were classified as for Experiment 1. Table 6 indicates the proportions of errors that were omissions, order errors, repetitions, intrusions, phonological errors, and unrelated errors for known and degraded words, combining across list lengths. As in the previous experiment, the patients' recall was characterized by abnormally frequent phonological errors. Unrelated errors also exceeded the normal range. In contrast to Experiment 1, none of the patients showed a difference in errors for known and degraded words [E.K.,  $\chi^2(5) = 2.39$ ; G.T.,  $\chi^2(5) = 2.12$ ; P.D.,  $\chi^2(5) = 2.52$ ; all n.s.]. This was largely because they made numerous phonological errors on both the known and the degraded words. Ten controls also showed no difference in error types between the known and the degraded words [ $\chi^2(5) = 1.80-7.83$ , all n.s.]. The difference in errors approached significance for one control [ $\chi^2(5) = 9.71$ ,  $p = .053$ ] and reached significance for another [ $\chi^2(5) = 15.91$ ,  $p < .01$ ].

The numbers of phonological and nonphonological errors occurring on the shortest two lengths (two and three words) were compared with those on the longest two lengths (four and five words), in order to determine whether list length affected the types of errors that were committed. P.D. was excluded from this analysis, since she had not been tested on the full complement of list lengths. The results were similar to those obtained in Experiment 1 (see Figures 2A and 2B). Phonological errors again appeared in large numbers on the shorter list lengths, whereas nonphonological errors (primarily omissions) increased sharply with length. E.K.'s phonological errors accounted for a greater proportion of the total errors on short, as compared with long, list lengths, for both known words [ $\chi^2(1) = 3.78$ ,  $p = .052$ ] and degraded words [ $\chi^2(1) = 14.19$ ,  $p < .001$ ], but the difference was more strongly significant for the degraded words, because there were

**Table 5**  
Recall of Frequency-Matched Known and Degraded Words  
Defined by Synonym Judgments (Experiment 2)

Patient	Words	List Length				
		2	3	4	5	7
E.K.	known	100.0	80.0*	37.5*	48.0*	–
	degraded	95.0	60.0*	42.5*	36.0*	–
E.K. controls (minimum)	known	–	100.0	80.0	66.0	34.3
	degraded	–	93.3	77.5	64.0	42.9
G.T.	known	85.0	80.0*	82.5	56.0	–
	degraded	70.0	53.3*	52.5*	28.0*	–
G.T. controls (minimum)	known	–	93.3	85.0	62.0	42.9
	degraded	–	86.7	77.5	62.0	47.1
P.D.	known	–	86.7	70.0	52.0	–
	degraded	–	90.0	67.5	46.0*	–
P.D. controls (minimum)	known	–	86.7	77.5	54.0	38.6
	degraded	–	90.0	67.5	52.0	44.3

Note—Figures indicate the percentages of items recalled in the correct order. \*Recall below minimum score obtained across all control participants on both known and degraded words.

**Table 6**  
**Errors on Frequency-Matched Known and Degraded Words Defined by Synonym Judgments (Experiment 2)**

Patient	Phonological		Unrelated		Omission		Order		Repetition		Intrusion	
	Known	Degraded	Known	Degraded	Known	Degraded	Known	Degraded	Known	Degraded	Known	Degraded
E.K.	.55*	.53*	.10*	.08*	.28	.35	.03	.03	.02	.00	.02	.01
G.T.	.66*	.59*	.11*	.15*	.24	.21	.00	.00	.00	.01	.00	.04
P.D.	.68*	.62*	.08*	.10*	.20	.24	.03	.02	.00	.02	.03	.00
All controls (maximum)	.14	.17	.06	.05	.74	.75	.32	.28	.18	.21	.37	.55

Note—The errors in each category are expressed as a proportion of the total number of errors across all list lengths. \*Patient scores were larger than the maximum observed for controls.

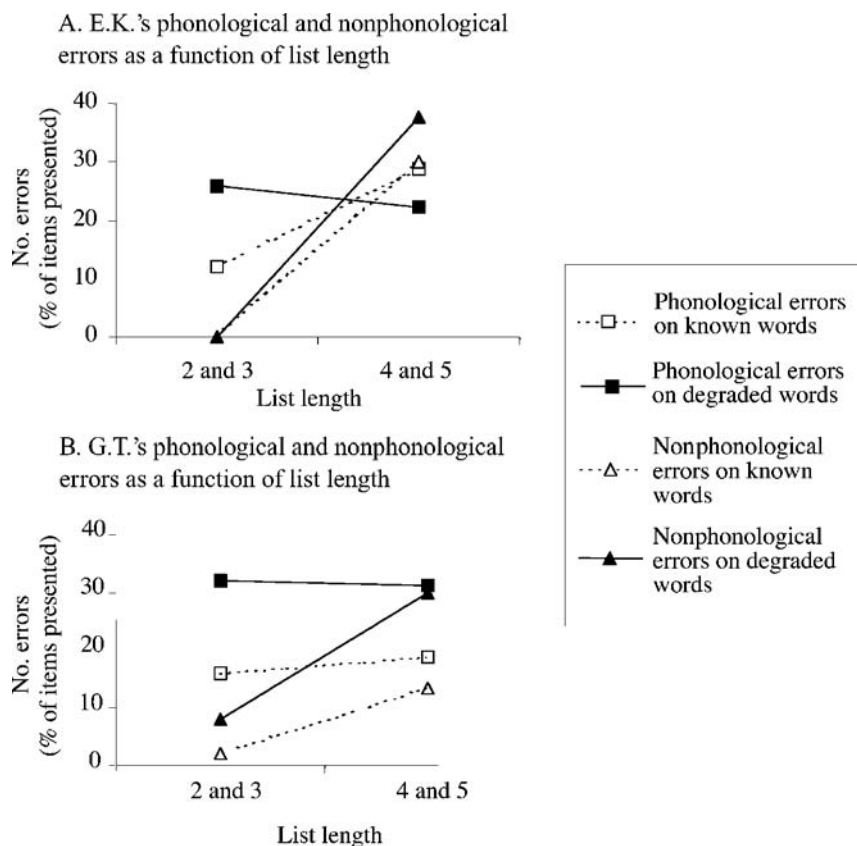
more errors. G.T. showed the same pattern for degraded words [ $\chi^2(1) = 3.99, p < .05$ ], but the difference did not reach significance for the known words [ $\chi^2(1) = 1.61, n.s.$ ].

### Discussion

As in Experiment 1, the patients' recall was impaired relative to control performance, with the exception of P.D., who showed normal recall accuracy despite her severe semantic impairments (although her error pattern was clearly abnormal). G.T. again showed a strong advantage in ISR for words that were relatively well understood, whereas E.K. and P.D. did not. There was, therefore, a

striking degree of consistency between the two experiments, with respect to which patients showed a known-degraded difference in recall accuracy, even though a very different method was used to select the known and the degraded words. It should be noted, however, that the results for E.K. were somewhat different: In Experiment 1, she showed significantly impaired recall of degraded but not known words, whereas in this experiment, her recall of both sets of items fell below the normal range.

As in the previous experiment, the patients made a large number of phonological errors in ISR, unlike the



**Figure 2. Phonological and nonphonological errors on known and degraded words defined by synonym judgment as a function of list length (Experiment 2).**

controls. In contrast to Experiment 1, however, these errors occurred frequently for both the known and the degraded words, and consequently, the patients did not show any error differences between the two conditions. Phonological errors may have occurred more frequently for the known words in this experiment because they were generally less imageable and familiar and may have been less well understood. List length did not affect the size of the known–degraded difference but did affect the proportion of phonological errors to other error types, as in Experiment 1. The errors on the shorter lists were predominantly phonological in nature. On longer lists, phonological errors remained frequent, but other types of errors (particularly omissions) became more common.

Experiments 1 and 2 suggest that certain patients consistently show substantial ISR differences between known and degraded words, whereas others do not. However, methodological factors could have been responsible for the failure to observe a known–degraded accuracy difference for E.K. and P.D., especially since they both showed the predicted difference in errors for one experiment. In the following experiments, we examined whether a significant difference between known and degraded words could be obtained for these patients in more favorable circumstances—for example, when the known and the degraded words were not matched for frequency and when set size was larger. Unfortunately, P.D. did not wish to participate in further research, and so we were able to investigate the role of these factors only in the recall of E.K., G.T., and M.K.

### EXPERIMENT 3

#### Known and Degraded Words Not Matched for Frequency

In the experiments above, the known and the degraded words were matched on an item-by-item basis for frequency, in order to ensure that superior known word recall

did not result from the normal frequency effect. However, as was mentioned in the introduction, frequency matching has a number of drawbacks, which may mask genuine known–degraded recall differences. It is particularly problematic that word frequency correlates with the known–degraded dimension. Since low-frequency items are also likely to be the most severely semantically degraded, frequency matching may limit the size of the semantic difference between known and degraded words. In addition, frequency matching restricts the number of experimental items that can be used (i.e., reduces the set size). In this experiment, the patients and the controls were compared on non–frequency-matched known and degraded words. The known words were higher in frequency and easier to recall than the degraded words for both the patients and the controls, but the known–degraded difference was expected to be much more substantial for the patients.

#### Method

Two sets of known and degraded words were selected, using naming and definitions, as in Experiment 1. The words were matched on an item-by-item basis for syllable length, but not for frequency, and consequently, the known words were substantially higher in frequency than the degraded words. Not matching the items for frequency allowed set size to be larger in this experiment (see Table A3 in the Appendix). Every participant was tested on four- and five-item lists. M.K. was additionally tested on three-item lists, and the controls were tested on seven-item lists. The lists were constructed following the method described in Experiment 1. There were 20 lists of known and degraded words at each length, and they were presented in blocks, using an ABBA design. Six age-matched control participants were tested on the material for each patient.

#### Results

Table 7 shows the percentage recall for the known and the degraded words. Both E.K. and G.T. recalled the known words at a normal level: Their recall was within the range of both the 6 controls tested on the same material, and all of the control data taken together (18 data

**Table 7**  
**Recall of Non-Frequency-Matched Known and Degraded Words (Experiment 3)**

Patient	Words	List Length			
		3	4	5	7
E.K.	known	–	91.3	78.0	–
	degraded	–	66.3*	48.0*	–
E.K. controls (minimum)	known	–	88.8	68.0	47.9
	degraded	–	91.3	71.0	52.1
G.T.	known	–	93.8	68.0	–
	degraded	–	78.8	46.0*	–
G.T. controls (minimum)	known	–	80.0	63.0	40.7
	degraded	–	78.8	55.0	43.6
M.K.	known	75.0*	60.0*	68.3*	–
	degraded	38.3*	36.7*	31.7*	–
M.K. controls (minimum)	known	97.0	84.0	81.0	50.0
	degraded	95.0	88.0	66.0	51.0

Note—Figures indicate the percentage of items recalled in the correct order. \*Score below range for each patient's controls.

points from 12 participants). M.K.'s recall of the known words was impaired, however, falling below the lowest scores obtained for controls. Recall of the degraded words fell substantially below the control range for all 3 patients.

Every patient recalled the known words significantly better than the degraded words when the data were combined across list lengths. E.K. showed a substantial advantage for the known words [ $t(71) = 5.92, p < .0001$ ], in contrast to the results in Experiments 1 and 2. The difference was also highly significant for G.T. [ $t(77) = 3.41, p < .001$ ] and M.K. [ $t(110) = 5.71, p < .0001$ ]. None of the controls showed a significant recall difference between known and degraded words [all  $t_s(114-158) < 1.24, n.s.$ ]. Therefore, the controls did not recall the known words better than the degraded words, even though they were higher in frequency.

The difference between the patients and the controls is further illuminated by a comparison of the size of the known-degraded recall difference. All of the data collected for the patients was included in this analysis, but only five- and seven-item lists were included for the controls (three- and four-item lists were excluded, since recall was close to ceiling, potentially reducing the size of the known-degraded difference). Since it was necessary to compare patients and controls on lists of different lengths, percentages of differences were contrasted. The known-degraded difference for all 18 controls was centered on zero (mean = 0.12, range = -6.3-5.4). The known-degraded differences for all 3 patients were much larger than the maximum difference observed in the controls (E.K., 38.9; G.T., 18.9; M.K., 24.2). The participants' errors followed the pattern reported for the previous experiments and so are not discussed in detail here.

## Discussion

In this experiment that examined known and degraded words not matched for frequency, E.K. showed a substantial recall advantage for the known words over the degraded words, in addition to G.T. and M.K., in contrast to the results of Experiments 1 and 2. Since the controls did not show this advantage, it seems unlikely that the patients' recall could be accounted for by the higher frequency of the known words or by any other difference between the two sets of words. Instead, the status of the words as known and degraded appeared to be crucial. The lack of frequency matching in this experiment allowed E.K.'s set size to be much larger than those in the previous experiments, and this difference in methodology may have accounted for the discrepancy in results. An alternative possibility, however, is that the known and degraded words varied in their extent of lexical-phonological degradation, as well as semantic degradation, given that word frequency effects are ascribed to the accessibility of lexical-phonological forms by some researchers (e.g., Jescheniak & Levelt, 1994). In the following experiments, the effect of set size on the recall of known and degraded words was examined more systematically, in

order to determine whether E.K. would show a significant known-degraded difference when set size was large, even for frequency-matched items.

## EXPERIMENT 4

### The Effect of Set Size on the Recall of Known and Degraded Words

The set sizes in Experiments 1 and 2 were small because of the constrained way in which the words were selected. In this experiment, the words from the two experiments were pooled in order to increase the set size.

## Method

The lists of four and five known and degraded words used in Experiments 1 and 2 were re-presented to E.K. and G.T., but were interspersed so that a list from Experiment 1 (based on naming and definitions) was followed by a list from Experiment 2 (based on synonym judgments). Consequently, a larger number of items occurred between repetitions of the same word. The lists in this new large set size condition were identical to those in Experiments 1 and 2, which constituted the small set size condition, and were presented in the same way. Six age-matched control participants were tested on the same material as the patients and also on lists of seven words.

## Results

**Recall accuracy.** Tables 8 and 9 show the percentages of words that were recalled by patients and controls in the large and small set size conditions, collapsing across four- and five-word lists. Table 8 shows the data for naming and definition items, and Table 9 shows the items based on synonym judgment. Since the patients were tested twice on the naming and definition words in the small set size condition (Experiment 1), the data in Table 8 are averaged across the two presentations. E.K. showed a significant difference between known and degraded

**Table 8**  
Effect of Set Size on the Recall of Known and Degraded Words Defined by Naming and Definitions (Experiment 4)

Patient	Set	Words	List Length	
			4+5	7
E.K.	large	known	73.3*	
		degraded	63.3*	
	small	known	70.6	
		degraded	64.4*	
E.K. controls (minimum)	large	known	78.9	50.0
		degraded	86.7	51.4
	small	known	68.9	48.6
		degraded	81.1	52.9
G.T.	large	known	84.4	
		degraded	62.2*	
	small	known	78.9	
		degraded	61.1*	
G.T. controls (minimum)	large	known	78.9	57.1
		degraded	82.2	52.9
	small	known	74.4	44.3
		degraded	75.6	50.0

Note—Figures indicate the percentages of items recalled in the correct order. \*Recall below minimum score obtained across all control participants on both known and degraded words.

**Table 9**  
**Effect of Set Size on the Recall of Known and Degraded Words**  
**Defined by Synonym Judgment (Experiment 4)**

Patient	Set	Words	List Length	
			4+5	7
E.K.	large	known	56.7*	
		degraded	31.1*	
	small	known	41.1*	
		degraded	38.9*	
E.K. controls (minimum)	large	known	70.0	41.4
		degraded	66.7	40.0
	small	known	73.3	40.0
		degraded	72.2	42.9
G.T.	large	known	67.8*	
		degraded	40.0*	
	small	known	67.8*	
		degraded	38.9*	
G.T. controls (minimum)	large	known	87.8	50.0
		degraded	80.0	44.3
	small	known	72.2	42.9
		degraded	70.0	47.1

Note—Figures indicate the percentages of items recalled in the correct order. \*Recall below minimum score obtained across all control participants on both known and degraded words.

words in the large set size condition [ $t(70) = 3.19, p < .01$ ], but not in the small set size condition [ $t(78) < 1$ ] supporting the notion that set size can affect the magnitude of known–degraded recall differences. A more detailed analysis revealed that, in the large set size condition, the known–degraded difference was significant for synonym judgment words [ $t(35) = 3.74, p < .001$ ] but not for naming and definition words [ $t(36) = 1.48, n.s.$ ]. Therefore, most of the effect of the set size manipulation was brought about by the synonym judgment words, which were recalled more poorly and with a greater proportion of phonological errors than the naming and definition words in Experiments 1 and 2. G.T. recalled the known words better than the degraded words regardless of whether set size was large [ $t(67) = 3.92, p < .001$ ] or small [ $t(76) = 4.64, p < .0001$ ]. This was the case for the words selected by means of naming and definitions [large set,  $t(29) = 2.54, p < .05$ ; small set,  $t(34) = 3.07, p < .01$ ], as well as by synonym judgments [large set,  $t(35) = 3.33, p < .01$ ; small set,  $t(38) = 3.86, p < .001$ ].

In contrast, none of the control participants showed a significant known–degraded recall difference for either naming and definition items or synonym judgments items in the large set size condition [all  $t_s(53-58) < 1.67, n.s.$ ]. This analysis combined scores on four-, five-, and seven-word lists. The performance of controls in the small set size condition was discussed for Experiments 1 and 2 and will not be considered here.

**Error analysis.** Table 10 shows the number of phonological and nonphonological errors (omission, order, repetition, intrusion, and unrelated errors) that occurred for the items selected according to naming and definitions and synonym judgment in the large and small set size conditions. Set size did not affect the balance of phonological to nonphonological errors produced by E.K., on either the naming and definitions items [known words,  $\chi^2(1) = 1.07, n.s.$ ; degraded words,  $\chi^2(1) < 1$ ] or the synonym judgment items [known and degraded words,  $\chi^2(1) < 1$ ]. Set size did not affect G.T.'s errors on the naming and definitions items [known and degraded words,  $\chi^2(1) < 1$ ] but did affect his errors on the synonym judgment items [known words,  $\chi^2(1) < 1$ ; degraded words,  $\chi^2(1) = 4.97, p < .05$ ]. A greater proportion of his errors were phonological in nature when the set size was large.

## Discussion

Set size affected the magnitude of the known–degraded recall difference for E.K., but not for G.T. E.K. recalled the known and the degraded words at an equivalent level in the small set size condition but showed a significant advantage for the known words over the degraded words in the large set size condition. E.K.'s recall of the degraded words defined by synonym judgments particularly seemed to benefit from small set sizes, perhaps because in Experiments 1 and 2 these words were more difficult to recall than those selected using naming and definitions. For G.T., a greater number of phonological errors occurred when set size was large, consistent with the notion that small set sizes improve recall by increasing the familiarity of degraded phonological forms. However, it remains unclear why set size failed to affect G.T.'s recall accuracy. One possibility is that the variation of set size was relatively subtle in this experiment. The manipula-

**Table 10**  
**Errors on Known and Degraded Words as a Function of Set Size (Experiment 4)**

Task	Patient	Set Size	Phonological		Nonphonological	
			Known	Degraded	Known	Degraded
Naming and definitions	E.K.	large	.04	.12	.23	.21
		small	.01	.16	.29	.19
	G.T.	large	.07	.21	.09	.17
		small	.06	.26	.14	.12
Synonym judgment	E.K.	large	.18	.31	.24	.37
		small	.29	.22	.30	.38
	G.T.	large	.23	.43	.09	.16
		small	.19	.31	.13	.30

Note—The errors in each category are expressed as a proportion of the number of items presented.

tion was applied only by changing the distance between repetitions of the same word and did not affect the total number of times each word was presented. In addition, overall set size was larger for E.K. than for G.T., so the distance between presentations of the same word may not have been large enough to produce an effect in G.T. We therefore examined the effect of set size in a second experiment that manipulated the total number of times each word was repeated and that kept set size equal for E.K. and G.T.

## EXPERIMENT 5

### A Second Look at the Effect of Set Size on the Recall of Known and Degraded Words

This experiment provided a replication of the previous set size findings using a rather different method.

#### Method

The methods used to select known and degraded words in previous experiments were relaxed in order to increase the number of words that were available for testing. Known words, selected using naming and definitions and synonym judgment, were supplemented with words that were produced correctly in fluency tasks. Degraded words were identified using the procedure described for Experiments 1 and 2. The known and the degraded words were matched for syllable length and frequency on an item-by-item basis, using data from Celex (Baayen et al., 1993) and the MRC psycholinguistic database (M. Coltheart, 1981). The characteristics of these words are described in Table A4 in the Appendix.

Four-item lists of known and degraded words were assembled at three set sizes: 36 words (each word was repeated twice), 12 words (each word was repeated 6 times), and 4 words (each word was repeated 18 times; i.e., each list contained the same four items in a different order). There were three different sets of 12 words, which together made up the complete set of 36, and there were three different sets of 4 words, which together made up the first set of 12 words. Therefore, any recall differences between the 36-word set and the 12-word set were likely to be the result of the number of times each word was repeated, and not of the words included in each set. There were 18 lists of known and degraded words in each set presented in blocks, using an ABBA design.

Testing took place over two sessions that were several weeks apart. On the first session, the 36-word set was tested, followed by the first set of 12 words and two 4-word sets. On the second session, the 36-word set was tested a second time in order to increase the amount of data available in that condition, followed by the remaining two sets of 12 words and the final set of 4 words. Three age-matched control participants were tested on the same lists as the patients and also on lists of 7 words (although they were tested on the 36-word set only once). The lists of 7 words necessarily included a larger number of repetitions of each word. For this reason, the controls were not tested on lists of 7 words in the 4-word set condition, since most of the items would have had to be presented twice in each list.

#### Results

**Recall accuracy.** Table 11 shows the percentage of known and degraded words that were recalled at each set size. The overall pattern was similar to that observed in Experiment 4. E.K. showed a more substantial recall difference between known and degraded words in the

**Table 11**  
Effect of Set Size on the Recall of Known and Degraded Words (Experiment 5)

Patient	Set Size	Words	List Length	
			4	7
E.K.	36	known	70.8*	
		degraded	50.7*	
	12	known	78.2*	
		degraded	70.4*	
	4	known	83.3*	
		degraded	71.8*	
E.K. controls (minimum)	36	known	90.3	46.8
		degraded	90.3	44.4
	12	known	90.7	49.5
		degraded	90.3	49.2
	4	known	95.4	–
		degraded	94.4	–
G.T.	36	known	77.1	
		degraded	61.1*	
	12	known	75.5*	
		degraded	54.2*	
	4	known	88.9*	
		degraded	62.5*	
G.T. controls (minimum)	36	known	75.0	35.7
		degraded	73.6	32.5
	12	known	80.6	49.5
		degraded	81.5	50.3
	4	known	89.8	–
		degraded	88.4	–

Note—Figures indicate the percentage of items recalled in the correct order. \*Recall below minimum score obtained across all control participants on both known and degraded words.

36-word set condition [ $t(66) = 3.52, p < .001$ ], as compared with the 4-word set condition [ $t(92) = 2.92, p < .01$ ], although the advantage for known over degraded words was still significant at the smallest set size. The known–degraded difference approached significance for the 12-word set [ $t(101) = 1.67, p < .1$ ]. As in the previous experiment, smaller set sizes facilitated E.K.'s recall of degraded words to a greater extent than her recall of known words. Set size did not significantly affect the recall of known words [36- vs. 12-word set,  $t(77) = 1.61$ ; 12- vs. 4-word set,  $t(98) = 1.38$ , both n.s.]. In contrast, degraded words were recalled better when the set size was smaller [36- vs. 12-word set,  $t(75) = 3.38, p < .01$ ; 12- vs. 4-word set,  $t(105) < 1$ ].

As in the previous experiment, G.T. showed a substantial recall difference between known and degraded words at every set size [36-word set,  $t(66) = 2.72, p < .001$ ; 12-word set,  $t(100) = 4.19, p < .0001$ ; 4-word set,  $t(102) = 5.99, p < .0001$ ]. The advantage G.T. showed for known words actually became greater as set size was decreased, in contrast with E.K., since decreases in set size enhanced the recall of known words more than recall of degraded words. Set size did not significantly affect the recall of degraded words [36- vs. 12-word set,  $t(78) = 1.14$ ; 12- vs. 4-word set,  $t(103) = 1.58$ , both n.s.]. However, known words were recalled better when the set size was smaller [36- vs. 12-word set,  $t(78) < 1$ ; 12- vs. 4-word set,  $t(104) = 3.21, p < .01$ ]. None of the control partic-

**Table 12**  
**Errors on Known and Degraded Words as a Function of Set Size (Experiment 5)**

Patient	Set Size	Phonological		Nonphonological	
		Known	Degraded	Known	Degraded
E.K.	36	.10*	.33*	.18	.15
	12	.06*	.16*	.15	.13
	4	.02*	.14*	.14*	.13*
G.T.	36	.17*	.26*	.07	.10
	12	.21*	.39*	.07	.05
	4	.07*	.31*	.05	.05
Controls (maximum): length = 4	36	.03	.04	.22	.22
	12	.02	.02	.19	.19
	4	.00	.00	.10	.11
Controls (maximum): length = 7	36	.06	.04	.64	.67
	12	.02	.02	.51	.51
	4	–	–	–	–

Note—The errors in each category are expressed as a proportion of the number of items presented. \*Patient scores were larger than the maximum observed for controls.

ipants showed a significant difference between the known and the degraded words at any set size, collapsing across list length [36-word set,  $t(69-70) < 1.14$ , n.s.; 12-word set,  $t(213-214) < 1$ ].

**Error analysis.** Table 12 shows the number of phonological and nonphonological errors (omission, order, repetition, intrusion, and unrelated errors) that were made at each set size. As in Experiments 1 and 2, a greater proportion of phonological errors occurred for the degraded words than for the known words. E.K. showed a significant known–degraded error difference for the 36-word set [ $\chi^2(1) = 10.34$ ,  $p < .01$ ], the 12-word set [ $\chi^2(1) = 6.92$ ,  $p < .01$ ], and the 4-word set [ $\chi^2(1) = 11.58$ ,  $p < .001$ ]. G.T. made a larger number of phonological errors on the known words, and consequently, the known–degraded error difference did not reach significance for the 36-word set [ $\chi^2(1) < 1$ ]. However, it did reach significance for the 12-word set [ $\chi^2(1) = 4.73$ ,  $p < .05$ ] and the 4-word set [ $\chi^2(1) = 6.48$ ,  $p < .05$ ].

For E.K., the pattern of errors changed smoothly over the three set sizes. Although there was no significant effect of set size on the pattern of errors when the known and the degraded words were considered separately [for all possible comparisons between set sizes,  $\chi^2(1) < 3.21$ , n.s.], the difference did reach significance when the errors made on known and degraded words were combined. A greater proportion of the errors were phonological for large, as compared with small, set sizes [36-word set vs. 4-word set,  $\chi^2(1) = 6.29$ ,  $p < .05$ ; 36-word set vs. 12-word set,  $\chi^2(1) = 3.60$ ,  $p = .06$ ; 12-word set vs. 4-word set,  $\chi^2(1) < 1$ ]. G.T. did not show this reduction in phonological errors with smaller set sizes. He showed no effects of set size on the balance of phonological to nonphonological errors for known words [for all comparisons,  $\chi^2(1) < 1.01$ , n.s.]. For degraded words, there were no significant differences in errors between the 36- and the 4-word sets [ $\chi^2(1) = 3.29$ , n.s.] or between the 12- and the 4-word sets [ $\chi^2(1) < 1$ ]. There were, however, more phonological errors in the 12-word set than in the 36-word set [ $\chi^2(1) = 5.61$ ,  $p < .05$ ].

## Discussion

Some interesting differences between E.K. and G.T. emerged in this experiment. E.K.'s recall of the degraded words improved as set size was decreased, and she made fewer phonological errors. She may have become more familiar with the phonological forms of the degraded words as they were repeated in the smaller set size conditions, and this would have increased the likelihood that their phonemes were produced in the correct configuration. Consequently, the magnitude of the recall difference between known and degraded words was smaller for more limited sets of words. In contrast, G.T. showed enhanced recall of known words when set size was small, but set size did not affect his recall of degraded words. As a result, the magnitude of the recall difference between known and degraded words was actually larger for more limited word sets.

These results suggest that the differences in the effect of set size observed for E.K. and G.T. in Experiment 4 did not occur because the manipulation of set size was relatively weak. Instead, these findings point to an effect of semantic knowledge on the degree to which repeatedly presenting the phonological forms of words enhances their immediate recall. In both patients, set size may have had the most substantial impact on those words that were partially comprehended. E.K.'s degraded words may have fallen into this category, whereas her known words were more semantically intact. In contrast, G.T. may have had very little understanding of his degraded words but some partial understanding of his known words, since his semantic impairments were more severe. Some support for this hypothesis can be provided by a closer look at the patients' definitions. Definitions were available for all of the words used in Experiment 4 and for the majority of the items used in Experiment 5. For the degraded words from Experiment 4, E.K. produced 2 correct definitions (for items selected according to synonym judgment) and 13 partially correct definitions. She failed to provide any correct information for 11 items. G.T. produced 2 correct definitions, 3 partially correct

definitions, and 18 entirely incorrect/no-response errors. This difference between the patients in the number of partially correct definitions was significant [ $\chi^2(2) = 7.91, p < .05$ ]. Similarly, for the degraded items from Experiment 5, E.K. produced 3 correct definitions, 18 partially correct definitions, and 12 entirely incorrect/no-response errors. G.T. produced no correct definitions, 14 partially correct definitions, and 21 entirely incorrect/no-response errors. Again, this difference was significant [ $\chi^2(2) = 5.47, p < .05$ ]. These findings are consistent with the suggestion that G.T. showed rather less of an effect of set size than did E.K. in degraded word recall because small set sizes improve the recall of partially degraded items to a greater extent than entirely degraded words. Snowden and Neary (2002) reached a similar conclusion in a study in which the relearning of verbal labels for pictures was examined. They found very poor learning of items that the patients apparently did not comprehend at all but substantial learning of items that the patients could not name but could demonstrate some knowledge of.

## COMBINED ANALYSIS

### Serial Position Effects in the Recall of Known and Degraded Words

Experiments 1–5 demonstrated, in line with previous studies (Knott et al., 1997, 2000; Patterson et al., 1994), that the likelihood of phonological disintegration in verbal STM is affected by the degree to which an item is semantically degraded. The analysis in this section combines the data from these experiments in order to examine the effect of semantic degradation on the shape of the serial position curve. There is considerable controversy about this topic. N. Martin and Saffran (1997) sought to account for the effect of lexical and semantic representations on phonological STM within Dell and O'Seaghda's (1992) interactive activation model of speech production and predicted that semantic factors should have their biggest impact on the early portions of the serial position curve. During ISR tasks, activation within this model spreads up from the phonological layer to lexical and semantic units and then back down again, allowing lexical and semantic activation to improve repetition accuracy. Since it takes a number of processing cycles for the activation to spread, the influence of semantics on ISR should be largest for the earliest presented words. Later portions of the serial position curve should be more heavily influenced by phonological factors, which occur more quickly.

There is some empirical support for the claim that semantic factors have the greatest influence in the initial portions of the serial position curve, both from studies of normal performance (Brooks & Watkins, 1990; Watkins & Watkins, 1977) and from neuropsychological studies (N. Martin & Saffran, 1990, 1997; R. C. Martin & Lesch, 1996; Saffran & Martin, 1990) showing that semantically impaired patients show a reduced primacy effect, whereas phonologically impaired patients show a re-

duced recency effect. N. Martin and Saffran (1997), for example, found an association in 15 aphasic patients between semantic impairment and the ability to recall both the first word of a two-word list and the initial phonemes of a single word. The literature is highly inconsistent, however. Hulme et al. (1997) found that the recall difference between high- and low-frequency words increased toward the *end* of the list, whereas Walker and Hulme (1999) obtained parallel serial position curves for concrete and abstract words. In addition, several neuropsychological studies have shown little difference in the shape of the serial position curve for relatively well-known and semantically degraded words (Forde & Humphreys, 2002; Knott et al., 1997). The semantically impaired patients in these previous studies showed primacy effects but negligible recency effects, contrary to the suggestion of N. Martin and Saffran (1997).

## Method

Serial position curves were derived for E.K., G.T., P.D., M.K., and their controls by amalgamating the data from Experiments 1–5 and the results of two similar experiments, not reported here. In every experiment, the patients had recalled matched lists of known and degraded words, and these words were examined separately. The largest amount of data was available for E.K. and G.T., and the smallest amount of data was available for P.D. For E.K. and G.T., there was considerably more data involving four-word lists than five-word lists. In total, 254 four-item lists and 110 five-item lists were available in the known and the degraded conditions. For their controls, 798 four-item lists, 420 five-item lists, and 636 seven-item lists were available, combining the data from 6 different participants. P.D. was tested on 20 four-item lists and 20 five-item lists. For her controls, 60 lists were available at each list length, combining the data from 3 participants. M.K. was tested on 50 four-item lists and 50 five-item lists. For M.K.'s controls, 180 lists were available at each list length, combining the data from 6 participants.

## Results

Figures 3A–3D show serial position curves for the 4 patients and their controls. The controls' serial position curves showed the standard pattern of markedly better recall at the beginning and end of the lists. For all 4 patients, there was no significant interaction between the known–degraded variable and serial position, on both four-item lists [E.K., M.K., and P.D.,  $\chi^2(3) < 1$ ; G.T.,  $\chi^2(3) = 2.76, n.s.$ ] and five-item lists [E.K., G.T., and P.D.,  $\chi^2(4) < 1$ ; M.K.,  $\chi^2(4) = 1.38, n.s.$ ]. E.K., G.T., and P.D. showed serial position effects that did not differ from those for their controls, for both known words [E.K.,  $\chi^2(4) = 1.34, n.s.$ ; G.T.,  $\chi^2(4) = 3.72, n.s.$ ; P.D.,  $\chi^2(4) < 1$ ] and degraded words [E.K.,  $\chi^2(4) = 1.27, n.s.$ ; G.T.,  $\chi^2(4) = 2.07, n.s.$ ; P.D.,  $\chi^2(4) < 1$ ]. These analyses examined recall on five-item lists. In contrast with the other patients, M.K. did not show normal effects of serial position, for either known words [ $\chi^2(4) = 15.99, p < .01$ ] or degraded words [ $\chi^2(4) = 13.29, p < .01$ ]. For Serial Positions 1–5, M.K.'s standardized residuals in these analyses were 2.3, –0.6, 1.4, –1.6, and –1.9 for known words and 1.5, 0.5, 1.5, –1.5, and –2.2 for degraded words. These values swing from positive to negative,



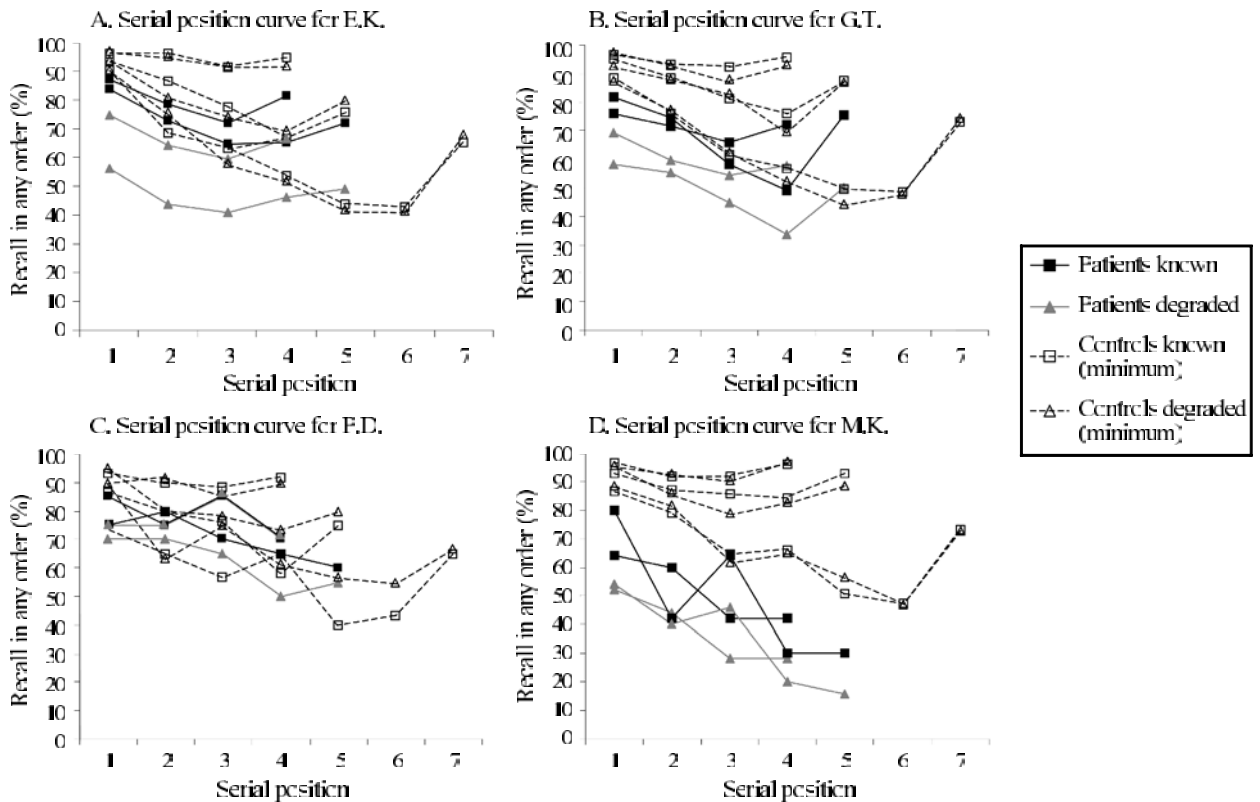


Figure 3. Serial position curves for known and degraded words.

suggesting that M.K.'s recall was poorer toward the end of the lists—that is, she did not show the standard recency effect.

## Discussion

E.K. and G.T. showed relatively typical serial position curves that were broadly parallel for known and degraded words. In contrast, M.K.'s recall fell steadily across serial positions for both known and degraded words, and she failed to show a normal recency effect. P.D. also showed little effect of recency in her recall, although for this patient, the effects of serial position did not diverge significantly from the normal pattern. Our findings are consistent with those in several other studies in which diminished recency effects in semantically impaired patients and comparable serial position curves for known and degraded words have been observed (Forde & Humphreys, 2002; Knott et al., 1997). We have found little evidence to support N. Martin and Saffran's (1997) contention that semantic factors have their biggest impact on the early portions of the serial position curve. According to N. Martin and Saffran (1997), semantic impairments should reduce the magnitude of the primacy effect but should leave the recency portion of the curve intact. For our patient M.K., the opposite appears to have been true.

How can our observations be reconciled with those in studies that have shown reduced primacy but intact recency

effects for semantically impaired patients (N. Martin & Saffran, 1990, 1997; Saffran & Martin, 1990) and for patients who have shown a specific impairment in retaining semantic information (R. C. Martin & Lesch, 1996)? There are several potentially important methodological differences between these studies and our own. N. Martin and Saffran's patients generally had substantial phonological impairments, whereas the patients in our study had relatively intact phonology; they almost never made phonological errors in picture naming or spontaneous speech and had normal digit spans. Consequently, N. Martin and Saffran's patients had much more severe deficits of verbal STM, and it was appropriate to test their ISR on very short lists (i.e., single items and pairs of items). In contrast, the patients in our study had much higher levels of ISR and were tested on considerably longer lists. This difference in list length may have important consequences for N. Martin and Saffran's (1997) prediction. Even if semantic activation lags behind phonological activation, as is predicted by Dell and O'Seaghdha's (1992) model, over the course of longer lists there may be time for semantic representations to become activated for all the list items, and consequently, semantic deficits may affect every portion of the serial position curve equally. In addition, given the requirement for serial order recall, the time lag between presentation and recall may have been similar across the different serial positions.

We still have to explain why the recency effect was diminished in the more severely semantically impaired patients. One possibility is that over the course of an ISR task, the phonological representations of the items to be recalled become noisy and unreliable (either because recall proceeds more slowly than presentation, causing more substantial phonological decay for the final list items, or because the recall of earlier list items interferes with the representation of subsequent items). In healthy individuals, the appropriate phonological activation may be sustained by lexical and semantic support, as was proposed by N. Martin and Saffran (1997). When this form of support is unavailable, however, as in the case of semantically degraded words, accurate phonological representations may not be maintained long enough to support recall of items at the end of the list (see Hulme et al., 1997, for a similar suggestion).

### GENERAL DISCUSSION

This series of experiments has explored the effect of various methodological factors on the ISR of relatively well-known and semantically degraded words in 4 patients with SD. Specifically, we varied the number of known and degraded words in the lists (set size), the method used to classify items as known and degraded, and the length of the lists to be recalled. In general, the patients recalled the known words more accurately than the degraded words and made a larger number of phonological errors on the degraded words, consistent with the notion that semantic representations play a major role in maintaining the phonological coherence of words in STM. In this discussion, we will examine the methodological variables that did and did not affect the magnitude of the recall difference between known and degraded words and will consider what these findings might suggest about the mechanism underlying the superior recall of known words.

In Experiments 1 and 2, which involved small sets of frequency-matched known and degraded words, two patients, G.T. and M.K., showed a significant advantage for the recall of known words over degraded words, but two other patients, E.K. and P.D., did not. In Experiment 3, the requirement that the known and the degraded words should be matched for frequency was relaxed, allowing set size to be substantially increased. Under these conditions, E.K. did recall significantly more known than degraded words, whereas the controls did not, suggesting that frequency differences per se could not account for the known–degraded difference. Experiments 4 and 5 more directly compared recall in large and small set size conditions and showed that when the same words were presented repeatedly, E.K.'s recall improved more substantially for degraded than for known words. E.K. made fewer phonological errors when set size was small, suggesting that her increased familiarity with the phonological forms of degraded words improved their coher-

ence in STM. Limited set size appeared to account for E.K.'s equivalent recall of known and degraded words in Experiments 1 and 2. Since many of the studies showing no recall difference between known and degraded words also used small set sizes (see Table 1), some of the discrepant results in the literature might be attributable to this factor.

Both E.K. and G.T. were included in Experiments 4 and 5, and set size did not affect them in exactly the same way. E.K. showed a substantial recall advantage for known over degraded words only when set size was large, whereas G.T. recalled the known words better than the degraded words at every set size. Set size particularly improved E.K.'s recall of degraded words. In contrast, for G.T., set size either had little impact on recall or enhanced recall of the known words to a greater extent than the recall of the degraded words. These findings point to an effect of semantic knowledge on the degree to which repeatedly presenting the phonological forms of words boosts their immediate recall. Set size might most strongly affect the recall of words that are partially semantically degraded but not completely forgotten. E.K.'s degraded words appear to have fallen into this category, whereas G.T.'s degraded words were more substantially affected. These results are consistent with those reported by Snowden and Neary (2002). They found that patients with SD are able to relearn the phonological forms of words that they still partly know but are much less able to relearn the phonological forms of words that have completely impoverished semantic representations.

How can the effect of set size on the recall of known and degraded words be explained? In healthy participants, recall is enhanced by small set sizes (V. Coltheart, 1993; Conrad, 1963), presumably because the words in small sets become more predictable as they are repeated. As a result, words from small sets should be easier to encode, retain, and recall correctly. Roodenrys and Quinlan (2000) proposed an explanation of the set size effect that was based on Hulme et al.'s (1997) *redintegration* theory. According to this account, noisy phonological traces in STM are automatically reconstructed from long-term lexical representations during the process of recall. The number of lexical candidates in the reconstructive process is reduced when set size is small, increasing the likelihood that the phonological trace will be restored accurately. Similar effects of set size might also be predicted by the interactive-activation account of N. Martin and Saffran (1997). The repeated presentation of items from limited sets might be expected to increase lexical and semantic activation for those items. The model predicts that this enhanced activation will constrain activity at the phonological level, increasing the likelihood that the target will be produced correctly.

These theories may also be able to account for the differential effect of set size on known and degraded words. Words that are severely degraded do not benefit from semantic support in ISR tasks, and consequently, their

phonological elements are not produced in the correct configuration. Set size would be expected to have little impact on these words with completely degraded semantics, since it is the lexical-semantic representations themselves that are thought to underpin the set size effect. Set size might also be expected to have little impact on the recall of very well known words, since these items will be adequately supported by intact semantic representations. Partially degraded words would be expected to derive the most support from small set sizes, since repeated presentation of these words might boost any residual lexical-semantic activation that still plays a role in maintaining phonological coherence in STM. Roodenrys and Quinlan (2000) found that set size affected the recall of lower frequency items only in normal participants, presumably because higher frequency items were adequately supported by their more accessible lexical representations. This result may be analogous to the effect of set size on the recall of well-known and partially degraded words in patients with SD.

We can also consider the impact of the methods used to select items as known and degraded. In Experiments 1 and 2, rather different semantic tests were employed to classify words as known and degraded, but in both experiments, G.T. and M.K. recalled the known words at a substantially higher level than the degraded words, whereas E.K. and P.D. did not. This consistency suggests that the methods used to select known and degraded items may not have a major impact on the magnitude of the known–degraded recall difference in terms of accuracy. Some clear differences did emerge, however, between the two sets of words. For the naming and definition words, phonological errors occurred more frequently for the degraded than for the known items, even for E.K. and P.D., who did not show an accuracy difference. There was no such error difference for the words selected according to synonym judgments, even for patients who showed an accuracy difference. The synonym judgment words were recalled more poorly than the naming and definition words, were more strongly affected by set size, and were characterized by frequent phonological errors affecting both known and degraded items. These findings suggest that the meanings of the synonym judgment words may have been generally more degraded than those of the naming and definition words.

Experiments 1 and 2 also examined the impact of list length on the recall of known and degraded words. The size of the known–degraded recall difference did not vary systematically with list length, and this factor had relatively little effect on the occurrence of phonological errors. In contrast, the number of nonphonological errors (predominantly omissions) rose steadily as list length was increased. We found no evidence to suggest that the coherence of items in STM was diminished in conditions of high phonological load. Phonological coherence was affected primarily by the type of material to be retained—that is, the status of the words as known or degraded—and not by the amount of material. This finding is consistent with the

suggestion that although semantic degradation impairs ISR performance in SD, the underlying phonological STM mechanism is intact, allowing these patients to show normal effects of phonological similarity and word length in ISR (see Jefferies et al., in press; Knott et al., 2000). Our patients were able to retain phonological representations of a relatively normal number of items in STM, although the coherence of these representations was weakened for degraded words, allowing phonological elements to migrate between the list items.

We have demonstrated that methodological factors, such as set size, can affect the magnitude of known–degraded recall differences in SD patients. In fact, almost all of the previous failures to find such differences can be accounted for by the use of small set sizes (see Table 1). The one exception is McCarthy and Warrington's (2001) patient M.N.A., who failed to show a recall difference between known and degraded words even though there were 30 items in each set. Like our patient P.D., M.N.A. had relatively normal recall accuracy for words that she understood poorly. It is important to note, however, that both P.D. and M.N.A. made frequent phonological errors in ISR, suggesting that their semantic degradation did affect the phonological coherence of items in STM.

Cases such as P.D. and M.N.A., who have good immediate recall despite severe semantic deficits, suggest that individual differences in dimensions other than semantic degradation may contribute to the degree of ISR impairment in this condition (although P.D.'s withdrawal from the study leaves some doubt about this issue). A potentially important factor is phonology: One might wonder whether G.T. and M.K., who showed the largest known–degraded recall differences, had phonological deficits in addition to their semantic impairments (see Jefferies et al., in press; Jefferies, Patterson, Bateman, & Lambon Ralph, 2004). If an interaction between phonology and semantics underpins verbal STM, the recall of degraded words might be expected to be particularly sensitive to phonological impairment, since these words should derive little support from semantics. Several authors have argued that additional phonological or lexical impairments may be required to produce phonological breakdown in ISR (Knott et al., 1997; McCarthy & Warrington, 2001), apparently challenging Patterson et al.'s (1994) claim that semantics plays a necessary and major role in constraining phonological activation in STM.

If the patients in this study had any phonological deficits, however, they must have been relatively subtle. The patients very rarely made phonological errors in spontaneous speech or picture naming. In addition, all 4 patients had normal digit span abilities, and E.K. and G.T. showed intact ISR for letters, although M.K. was mildly impaired (P.D. was not included in this study). Moreover, E.K., G.T., and M.K. showed effects of phonological similarity in ISR that were within the normal range (again, no data are available for P.D.; Jefferies, Patterson, Jones, Bateman, & Lambon Ralph, 2004), suggesting that their verbal STM abilities relied heavily on a phonological

code, as in healthy participants. Nevertheless, on the basis of these findings, we are unable to rule out the possibility that some degree of phonological impairment contributed to the more robust and consistent effects of the known–degraded manipulation observed for G.T. and M.K. On a range of phonological-processing tests (e.g., minimal pairs, phoneme segmentation, rhyme judgment and production, and nonword repetition and recall), E.K. showed good performance. In contrast, G.T. displayed some weaknesses in phoneme segmentation and nonword recall (Jefferies et al., in press; P.D. and M.K. were not included in this experiment). In addition, at a late stage of the condition, M.K. began to show phonological errors in naming and immediate single-word repetition that might have been indicative of phonological problems (Jefferies, Patterson, Bateman, & Lambon Ralph, 2004). It should be noted, however, that theories that posit substantial interaction between semantics and phonology (e.g., Patterson et al., 1994) might predict an effect of severe semantic degradation on phonological processing as well as on ISR, and consequently, these findings do not unequivocally demonstrate the presence of an independent phonological impairment for G.T. and M.K.

The issue of whether known–degraded recall differences can emerge in the absence of subtle phonological processing deficits is an important empirical question that we have sought to address in another study (Jefferies et al., in press). In this investigation, phonology and verbal STM were examined in a group of 6 SD patients (including E.K. and G.T.). Although it was not possible to rule out phonological impairment in some of them, substantial known–degraded recall differences emerged even for patients who performed well on all the tests of phonological processing (e.g., minimal pairs, phoneme segmentation, rhyme judgment and production, and nonword repetition and recall), supporting the view that semantics makes a major contribution to verbal STM.

If this conclusion is correct, how can we account for the failure of P.D. and M.N.A. to show a significant recall advantage for relatively well-understood words? Moreover, do these cases seriously challenge the notion of a critical semantic contribution to verbal STM, given that they appear to display intact ISR despite severe semantic impairments? Not necessarily. One possibility consistent with Patterson et al.'s (1994) semantic binding hypothesis is that these individuals have sophisticated phonological systems that can maintain the phonological elements of words in the correct configuration with minimal support from semantics. Indeed, both P.D. and M.N.A. had exceptional digit spans of around eight items, and P.D.'s digit recall exceeded the maximum achieved by controls on some list lengths (Jefferies, Patterson, Jones, Bateman, & Lambon Ralph, 2004), suggesting that she did have superior phonological abilities (unfortunately, however, P.D.'s withdrawal from the study prevented us from obtaining a more detailed assessment of her phonological capacities).

An analogy can be drawn from the domain of reading, where it has been argued that semantics plays a major role in translating between orthography and phonology (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996). According to this account, semantically impaired patients should be poor at reading low-frequency words with atypical spelling-to-sound correspondences, and indeed many semantically impaired patients do show this pattern (K. S. Graham, Hodges, & Patterson, 1994; Patterson & Hodges, 1992). There are, however, reports of a few semantically impaired patients who were normal at reading low-frequency exception words (Cipolotti & Warrington, 1995; Lambon Ralph, Ellis, & Franklin, 1995). According to Plaut (1997), these cases had highly developed phonological pathways that mastered the problem of learning to read without the usual level of semantic involvement. Similarly, P.D. and M.N.A. might have had good phonological skills as their verbal STM abilities developed, causing their mature phonological systems to be less reliant on semantic support. By this view, although verbal STM is underpinned by automatic interactions between phonology and semantics in the mature system, interaction between these representations during development determines the eventual balance of labor in ISR tasks. If for a particular individual, well-developed phonological representations are able to map between acoustic and articulatory codes with little contribution from semantics, the mature phonological system will remain relatively independent of the semantic system. According to this conjecture, patients who fail to show an ISR difference between known and degraded words should have excellent phonological skills; it will clearly be important in future research to establish whether this is the case.

There appears to be no good reason for rejecting the view that semantics makes an important contribution to the phonological coherence of items in STM. The majority of the failures to find ISR differences between known and degraded words can be accounted for by methodological factors. Moreover, patients who show smaller ISR impairments than might be expected given their level of semantic degradation do not necessitate functional independence between phonology and semantics. Instead, these patients may represent one end of a spectrum of phonological and semantic support for ISR. Our results are most consistent with the view that verbal STM is underpinned by interactions between the full range of representations, including phonology and semantics, which underlie language processing.

## REFERENCES

- BAAYEN, R. H., PIEPENBROCK, R., & VAN RIJN, H. (1993). *The CELEX Lexical Database* [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.
- BADDELEY, A. D., & HITCH, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-90). New York: Academic Press.

- BOURASSA, D. C., & BESNER, D. (1994). Beyond the articulatory loop: A semantic contribution to serial order recall of subspan lists. *Psychonomic Bulletin & Review*, **1**, 122-125.
- BOZEAT, S., LAMBON RALPH, M. A., PATTERSON, K., & HODGES, J. R. (2002). The influence of personal familiarity and context on object use in semantic dementia. *Neurocase*, **8**, 127-134.
- BROOKS, J. O., III, & WATKINS, M. J. (1990). Further evidence of the intricacy of memory span. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 1134-1141.
- CAZA, N., BELLEVILLE, S., & GILBERT, B. (2002). How loss of meaning with preservation of phonological word form affects immediate serial recall performance: A linguistic account. *Neurocase*, **8**, 225-273.
- CIPOLOTTI, L., & WARRINGTON, E. K. (1995). Semantic memory and reading abilities: A case report. *Journal of the International Neuropsychological Society*, **1**, 104-110.
- COLTHEART, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, **33A**, 497-505.
- COLTHEART, V. (1993). Effects of phonological similarity and concurrent irrelevant articulation on short-term memory recall of repeated and novel words lists. *Memory & Cognition*, **21**, 539-545.
- CONRAD, R. (1963). Acoustic confusion and memory span for words. *Nature*, **197**, 1029-1030.
- CONRAD, R. (1964). Acoustic confusion in immediate memory. *British Journal of Psychology*, **55**, 75-84.
- CONRAD, R., & HULL, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology*, **55**, 429-432.
- DELL, G. S., & O'SEAGHDA, P. G. (1992). Stages of lexical access in language production. *Cognition*, **42**, 287-314.
- EVANS, J. J., HEGGS, A. J., ANTOUN, N., & HODGES, J. R. (1995). Progressive prosopagnosia associated with selective right temporal lobe atrophy: A new syndrome? *Brain*, **118**, 1-13.
- FOLSTEIN, M. F., FOLSTEIN, S. E., & MCHUGH, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, **12**, 189-198.
- FORDE, E. M. E., & HUMPHREYS, G. W. (2002). The role of semantic knowledge in short-term memory. *Neurocase*, **8**, 13-27.
- FUNNELL, E. (1995). Objects and properties: A study of the breakdown of semantic memory. *Memory*, **3**, 497-518.
- FUNNELL, E. (1996). Response biases in oral reading: An account of the co-occurrence of surface dyslexia and semantic dementia. *Quarterly Journal of Experimental Psychology*, **49A**, 417-446.
- GRAHAM, K. S., HODGES, J. R., & PATTERSON, K. (1994). The relationship between comprehension and oral reading in progressive fluent aphasia. *Neuropsychologia*, **32**, 299-316.
- GRAHAM, N. L., PATTERSON, K., & HODGES, J. R. (2001). The emergence of jargon in progressive fluent dysgraphia: The widening gap between target and response. *Cognitive Neuropsychology*, **18**, 343-361.
- HODGES, J. R., PATTERSON, K., OXBURY, S., & FUNNELL, E. (1992). Semantic dementia: Progressive fluent aphasia with temporal lobe atrophy. *Brain*, **115**, 1783-1806.
- HOWARD, D., & PATTERSON, K. (1992). *Pyramids and Palm Trees: A test of semantic access from pictures and words*. Bury St. Edmunds, U.K.: Thames Valley Test Company.
- HULME, C., MAUGHAN, S., & BROWN, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory & Language*, **30**, 685-701.
- HULME, C., ROODENRYS, S., BROWN, G., & MERCER, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology*, **86**, 527-536.
- HULME, C., ROODENRYS, S., SCHWEICKERT, R., BROWN, G. D. A., MARTIN, S., & STUART, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a reintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 1217-1232.
- JEFFERIES, E., JONES, R., BATEMAN, D., & LAMBON RALPH, M. A. (in press). A semantic contribution to nonword recall? Evidence for intact phonological processes in semantic dementia. *Neuropsychology*.
- JEFFERIES, E., PATTERSON, K., BATEMAN, D., & LAMBON RALPH, M. A. (2004). *The natural history of late-stage "pure" semantic dementia*. Manuscript in preparation.
- JEFFERIES, E., PATTERSON, K., JONES, R. W., BATEMAN, D., & LAMBON RALPH, M. A. (2004). A category-specific advantage for numbers in verbal short-term memory: Evidence from semantic dementia. *Neuropsychologia*, **42**, 639-660.
- JESCHENIAK, J. D., & LEVELT, W. J. M. (1994). Word-frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 824-843.
- KNOTT, R., PATTERSON, K., & HODGES, J. R. (1997). Lexical and semantic binding effects in short-term memory: Evidence from semantic dementia. *Cognitive Neuropsychology*, **14**, 1165-1216.
- KNOTT, R., PATTERSON, K., & HODGES, J. R. (2000). The role of speech production in auditory-verbal short-term memory: Evidence from progressive fluent aphasia. *Neuropsychologia*, **38**, 125-142.
- KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- LAMBON RALPH, M. A., ELLIS, A. W., & FRANKLIN, S. (1995). Semantic loss without surface dyslexia. *Neurocase*, **1**, 363-369.
- LAMBON RALPH, M. A., & HOWARD, D. (2000). Gogi aphasia or semantic dementia? Simulating and assessing poor verbal comprehension in a case of progressive fluent aphasia. *Cognitive Neuropsychology*, **17**, 437-465.
- LEZAK, M. (1976). *Neuropsychological assessment*. New York: Oxford University Press.
- MARTIN, N., & SAFFRAN, E. M. (1990). Repetition and verbal STM in transcortical sensory aphasia: A case study. *Brain & Language*, **39**, 254-288.
- MARTIN, N., & SAFFRAN, E. M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology*, **14**, 641-682.
- MARTIN, R. C., & LESCH, M. F. (1996). Associations and dissociations between language impairment and list recall: Implications for models of STM. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 149-178). Hove, U.K.: Psychology Press.
- MARTIN, R. C., LESCH, M. F., & BARTHA, M. C. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory & Language*, **41**, 3-29.
- MARTIN, R. C., SHELTON, J., & YAFFEE, L. S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory & Language*, **33**, 83-111.
- MCCARTHY, R. A., & WARRINGTON, E. K. (1987). The double dissociation of short-term memory for lists and sentences: Evidence from aphasia. *Brain*, **110**, 1545-1563.
- MCCARTHY, R. A., & WARRINGTON, E. K. (2001). Repeating without semantics: Surface dysphasia? *Neurocase*, **7**, 77-87.
- PATTERSON, K., GRAHAM, N., & HODGES, J. R. (1994). The impact of semantic memory loss on phonological representations. *Journal of Cognitive Neuroscience*, **6**, 57-69.
- PATTERSON, K., & HODGES, J. R. (1992). Deterioration of word meaning: Implications for reading. *Neuropsychologia*, **30**, 1025-1040.
- PLAUT, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language & Cognitive Processes*, **12**, 765-805.
- PLAUT, D. C., MCCLELLAND, J. L., SEIDENBERG, M., & PATTERSON, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, **103**, 56-115.
- POIRIER, M., & SAINT AUBIN, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *Quarterly Journal of Experimental Psychology*, **48A**, 384-404.
- RAVEN, J. C. (1962). *Coloured progressive matrices sets A, AB, B*. London: H. K. Lewis.
- ROODENRYS, S., & QUINLAN, P. T. (2000). The effects of stimulus set size and word frequency on verbal serial recall. *Memory*, **8**, 71-78.
- SAFFRAN, E. M., & MARTIN, N. (1990). Neuropsychological evidence for lexical involvement in short-term memory. In G. Vallar & T. Shal-

- lice (Eds.), *Neuropsychological impairments of short-term memory* (pp. 145-166). Cambridge: Cambridge University Press.
- SNOWDEN, J. S., GOULDING, P. J., & NEARY, D. (1989). Semantic dementia: A form of circumscribed cerebral atrophy. *Behavioural Neurology*, **2**, 167-182.
- SNOWDEN, J. S., & NEARY, D. (2002). Relearning of verbal labels in semantic dementia. *Neuropsychologia*, **40**, 1715-1728.
- SNOWDEN, J. S., NEARY, D., & MANN, D. M. A. (EDS.) (1996). *Frontotemporal lobar degeneration: Frontotemporal dementia, progressive aphasia, semantic dementia*. London: Churchill Livingstone.
- TREIMAN, R., & DANIS, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 145-152.
- WALKER, I., & HULME, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 1256-1271.
- WARRINGTON, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology*, **27**, 635-657.
- WARRINGTON, E. K., & JAMES, M. (1991). *The visual object and space perception battery*. Bury St. Edmunds, U.K.: Thames Valley Test Company.
- WATKINS, O. C., & WATKINS, M. J. (1977). Serial recall and the modality effect: Effects of word frequency. *Journal of Experimental Psychology: Human Learning & Memory*, **3**, 712-718.
- WECHSLER, D. (1987). *Wechsler Memory Scale-Revised*. New York: Psychological Corporation.
- WECHSLER, D. (1997). *Wechsler Memory Scale-Third Edition*. London: Psychological Corporation.

## APPENDIX

**Table A1**  
**Properties of the Known and Degraded Words Selected for Each Patient in Experiment 1**

Patient	Words	Set Size	Lemma Frequency (Celex)	Frequency (Kučera & Francis, 1967)	Imageability	Syllable Length
E.K.	known	11	18.6	15.7* (18%)	590* (27%)	2.2
	degraded		19.1	15.2* (18%)	591	1.7
G.T.	known	9	18.0	14.6* (11%)	560* (22%)	2.0
	degraded		18.3	14.0* (22%)	584* (33%)	2.0
P.D.	known	6	59.9	28.4	587	1.6
	degraded		51.7	27.8	614	1.6
M.K.	known	7	85.1	52.0	602	1.1
	degraded		76.1	54.1	587* (14%)	1.1

Note—Imageability values are taken from the MRC psycholinguistic database (M. Coltheart, 1981) and range from 100 to 700. \*Values were unavailable for some items (figure in parentheses indicates the percentage of items affected).

**Table A2**  
**Properties of the Known and Degraded Words Selected for Each Patient in Experiment 2**

Patient	Words	Set Size	Lemma Frequency (Celex)	Frequency (Kučera & Francis, 1967)	Imageability	Syllable Length
E.K.	known	15	39.3	30.5	419	2.3
	degraded		37.8	38.6	390	2.4
G.T.	known	14	54.9	48.5	434	2.1
	degraded		53.3	50.7	405	2.2
P.D.	known	14	43.6	43.1	480	1.9
	degraded		44.6	37.7	447	2.2

Note—Imageability values are taken from the MRC psycholinguistic database (M. Coltheart, 1981) and range from 100 to 700.

**Table A3**  
**Properties of the Known and Degraded Words Selected for Each Patient in Experiment 3**

Patient	Words	Set Size	Lemma Frequency (Celex)	Frequency (Kučera & Francis, 1967)	Imageability	Syllable Length
E.K.	known	23	95.7	64.8* (9%)	587* (17%)	1.7
	degraded		107.7	30.0* (13%)	592* (26%)	1.7
G.T.	known	22	138.5	113.1* (14%)	584* (23%)	1.6
	degraded		20.7	14.1* (9%)	591* (23%)	1.6
M.K.	known	26	169.3	168.9	599* (4%)	1.3
	degraded		24.2	33.2* (4%)	589* (15%)	1.3

Note—Imageability values are taken from the MRC psycholinguistic database (M. Coltheart, 1981) and range from 100 to 700. \*Values were unavailable for some items (figure in parentheses indicates the percentage of items affected).

**Table A4**  
**Properties of the Known and Degraded Words Selected for Each Patient in Experiment 5**

Patient	Words	Lemma Frequency (Celex)	Frequency (Kučera & Francis, 1967)	Imageability	Syllable Length
E.K.	known	23.5	22.9* (14%)	559	2.1
	degraded	23.9	23.5* (14%)	492	2.1
G.T.	known	34.8	36.1* (17%)	508* (19%)	1.8
	degraded	32.1	31.6* (11%)	487* (22%)	1.8

Note—Imageability values are taken from the MRC psycholinguistic database (M. Coltheart, 1981) and range from 100 to 700. \*Values were unavailable for some items (figure in parentheses indicates the percentage of items affected).