The Levels of Processing Effect on Memory for Words

Mike Griffiths

First Year B Sc (Hons) Psychology,
Goldsmiths College, New Cross, London SE14 6NW.

Laboratory Group E
Laboratory Report No 8: experiment undertaken on 27 February 2002

Abstract

Craik and Lockhart (1972, in Eysenck and Keane, 2000) have proposed that the ability to remember words depends on how deeply they were encoded when encountered. In a similar fashion to Craik and Tulving (1975, in Baddely, 1997), participants in the present experiment were given words and asked to make physical, acoustic or semantic judgements about them. In a subsequent test, participants recognised more words in the conditions which were considered to involve deeper coding. They also recognised more words when the answer to their judgement had been positive. The results were interpreted being consistent with the current understanding of memory processes.

Introduction

Most theories of memory (e.g. Atkinson and Shiffrin, 1968, in Eysenck and Keane, 2000) have described it in terms of having three types of store: sensory, short term and long term. It is held that information must flow through each stage to reach the next. The distinction between short and long term store is analogous to that proposed by William James (1890, in Eysenck and Keane, 2000). This primary or short term memory is generally considered to contain the information which is currently in consciousness, perhaps being used in some mental task or simply repeated in an attempt to remember it.

Craik and Lockhart (1972, in Eysenck and Keane, 2000) put forward a different perspective for understanding memory. They claimed that the information stored in long term memory is determined by attentional and perceptual processes at the time of learning. A key concept was that there are various levels of processing that can be undertaken on a stimulus, from physical analysis (e.g. detecting individual letters) to deep or 'semantic' analysis. Craik (1973, in Eysenck and Keane, 2000, page 165) defined depth as "the meaningfulness extracted from the stimulus rather than … the number of analyses carried out on it". Thus, Craik and Lockhart assumed that the level of processing of a stimulus determined how well it was remembered.

In support of their approach, Craik and Tulving (1975, in Baddely, 1997) presented participants with sequences of words, and asked them to carry out tasks which were supposed to require different depths of processing. In order of presumed depth, they were asked to either (a) say whether the words were in upper or lowercase letters; (b) ask whether it rhymed with another; (c) decide whether it fitted in a given sentence. Participants' memory was better for the more deeply-encoded words. In a subsequent (surprise) test, participants did indeed recognise more of the deeply-processed words. They also remembered more words for which the answer to the question had been 'yes' than 'no'.

As Glanzer (1977) has pointed out, Craik and Lockhart's perspective is not really an alternative to the muliti-store model, but could simply be an account of how information is transferred from short term store to long term store. Eysenck and Keane (2000) point out that an important departure by Craik and Lockhart from the traditional model is that they distinguished between 'maintenance rehearsal' – repeating analyses that have previously been carried out - and 'elaborative rehearsal' – deeper analysis of the material. Eysenck and Keane (2000) cite Craik and Lockhart (1972) as having claimed that maintenance rehearsal provides no leaning at all, but say that they themselves consider this to be an overstatement of the position.

Eysenck and Keane (2000) identify at least two, possibly associated, dimensions that affect transfer into long term memory. The first, from Craik and Tulving (1975) is elaboration of processing – the amount of processing of a given kind. Thus, when participants were asked to say whether words fitted into a sentence, their recall was better the more complex the sentence. The second is distinctiveness of processing. Eysenck (1979, in Eysenck and Keane 2000) argued for the importance of this; in support they found that pronouncing words such as 'comb' as if the 'b' was not silent greatly increased their recognition.

Many of the experiments in this field have involved incidental learning, in which the participant is told only that they are to carry out a processing task (i.e. to answer questions about a word) and is not told that they be given a subsequent test of their memory for it. Gregg (1975) has pointed out the difficulty of maintaining the necessary deception to study this phenomenon, but nonetheless it does appear that it can be demonstrated. As he points out, it is somewhat mysterious how participants can retrieve words when they are unexpectedly required to do so and have hence set up no plan for it.

Baddely (1997) has analysed a number of problems with the Levels of Processing concept. The first is that there appears to be no independent method of measuring the 'depth of processing': a form of processing is presumed to be deeper if it leads to better learning. With only this circular definition, the approach may appear intuitively helpful but its scientific validity is limited. He also notes that there has been criticism of the approach's concentration on encoding, rather than including a consideration of the retrieval conditions. Thus Morris, Bransford and Franks (1977, in Baddely, 1997) gave participants words (e.g. cat) and asked them to make either a rhyme judgement (e.g. does it rhyme with mat) or semantic judgements (e.g. does it have a tail). When asked to recognise words (as in most experiments which have been performed in this field) the participants performed better with the semantically-processed words: but when asked whether there had been words which rhymed with listed words they performed better with the rhyming-processed words.

The present experiment was an attempt to replicate the findings of Craik and Tulving (1975, in Baddely, 1997) in relation to incidental learning of words which were processed in different ways.

The hypotheses were that:

1. Semantically-processed words would be recalled better than acoustically-processed words, and acoustically-processed words would be recalled better than physically-processed words;
2. Words would be recalled better if the answer to the processing task was 'yes' than if the answer was 'no'.

Method

Design

The experiment had a within-subjects design. The independent variables were the type of processing required to answer questions about the stimuli, and whether the answer to the question was 'yes' or 'no'. The dependent variable was the number of words recognised on a subsequent test.

Participants

The participants were an opportunity sample of 20 first year undergraduate psychology students, 5 male and 15 female, participating as part of a course requirement. They ranged in age from approximately 18-50 years. Some participants had English as their second language.

Materials and Stimuli

48 common English words were used, presented one at a time, together with a statement about them which could be true or false. The questions were assigned pseudo-randomly such that there were 16 statements in each condition, of which half were true. Questions and statements were in capital letters. Examples for each level of processing are given in Table 1:

Table 1: Example Stimuli

Question	Word	Correct Answer	Type of Processing
THE WORD HAS FIVE LETTERS	INSANE	N	Visual
THE WORD RHYMES WITH OCEAN	MOTION	Y	Acoustic
THE WORD FIRST IN THE SENTENCE "----- AEROPLANES IS VERY EXHILARATING"	FLYING	Y	Semantic

There were a further six word-sentence pairs for practice trials. 48 similar words were used as distractors in the recognition task.

A fixation point in the form of a cross (+) was also used.

Personal Computers were used to display the stimuli.

Procedure

Participants were given instructions for the encoding task in writing and on-screen, and the opportunity to ask questions. They were told that there would be a further judgment task, but not what it was.

Participants were told that the questions would be one of three types:

1. whether a statement about the length of a word was true;
2. whether it rhymed with another word;
3. whether it made sense to fit the word into a blank sentence.

They were given example questions to explain these types of question.

In the encoding task, each statement was presented for an adequate time to be read, then a fixation point was presented below and to the right of it. Soon afterwards, the fixation point was replaced by the target word[1]. Participants indicated whether the statement was true or false with respect to the target word, by pressing 'T' or 'F' accordingly. There were told to respond as quickly as possible whilst keeping errors to a minimum. There were six practice trials followed by 48 test trials.

After the coding task, participants given the recognition task. They were presented with on-screen instructions which told them that they would see an equal number of words which had and had not been targets in the previous task. They were to press

[1] These times were standardised but their length was not available at the time of writing this report.

'Y' if they had seen the word before and 'N' if they had not. The computer then presented them with the 48 words that they had previously seen, randomly mixed with 48 new words. Unlimited time was allowed for participants' responses.

When this was completed, participants were given their results in the format shown in Appendix 1. The number of words correctly recalled for each member of the group were then pooled.

Results

The detailed results are given in Appendix 2. The Means and Standard Deviations for each condition are shown in Table 2, and the Means in Figure 1.

Table 2: Number of Words Correctly Recognised

	True Statement			False Statement		
	Visual	Acoustic	Semantic	Visual	Acoustic	Semantic
Mean	**4 .2**	**5 .9**	**6 .5**	**3 .4**	**4 .3**	**5 .5**
S.D.	**1 .8**	**1 .9**	**1 .1**	**1 .4**	**1 .8**	**1 .7**
% of possible words correct	52%	74%	81%	43%	54%	68%

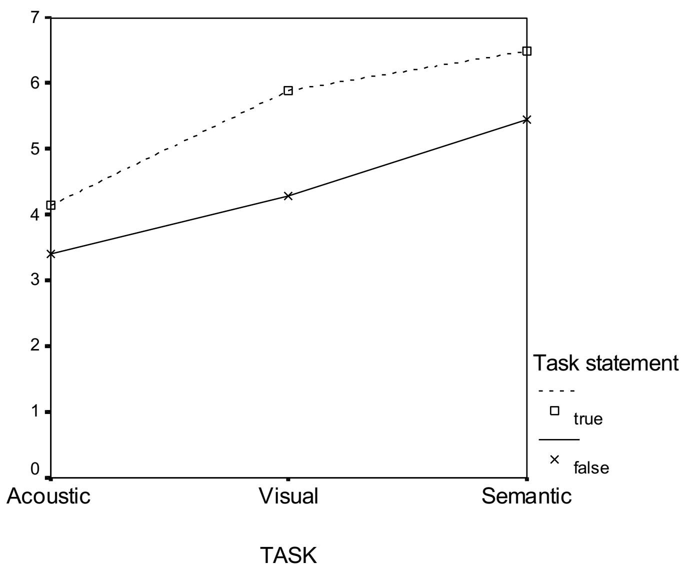

Figure 1: Mean number of words recalled

A two-factor within-subjects ANOVA (Appendix 3[2]) shows that there were overall significant differences in the number of words recognised, both according to the types of encoding [$F(2,38) = 39.246$, $p < 0.001$] and to whether the original statement was true or false [($F(1,19) = 24.568$, $p < 0.001$]. There were no significant interaction effects [$F(2,38) = 1.071$, N.S.].

[2] As an alternative, two one-way ANOVAs are appended (Appendix 4) with an interpretation based on them (Appendix 5).

Bonferroni-corrected pairwise comparisons show that there are significant differences between each possible pair of coding conditions (for visual and acoustic $p < 0.001$; for visual and semantic $p < 0.001$; for acoustic and semantic $p = 0.006$).

The experimental hypotheses were therefore supported.

<u>Discussion</u>

The experiment replicated the findings of Craik and Tulving (1975, in Baddely, 1997). However, it is still subject to the same criticisms, for example that the definition of 'depth of processing' is circular. It has therefore not added anything to the understanding of the 'depth of processing' concept.

As Morris, Bransford and Franks' (1977, in Baddely, 1997) experiment highlights, the independent variable in the present experiment may manipulate factors which are not relevant to the depth of processing (e.g. requiring auditory rehearsal in the case of the rhyming task). From this point of view, experiments such as that of Craik and Tulving (1975, in Eysenck and Keane, 2000), where all the tasks involve sentences but some are more complicated or distinctive, are more convincing evidence of the 'levels of processing' concept.

It is possible that the effect found was simply due to participants taking longer to answer the questions in the different conditions. This could have been tested had the times taken been recorded and analysed. It might have been possible to control for this possibility, by giving participants a limited time regardless of condition, but in any case participants would stop processing the stimulus once their judgement was made, and if the memory test were genuinely unexpected they would presumably stop paying any attention to it at all. This could perhaps be overcome by giving participants multiple tasks on each stimulus until a standard time had elapsed (e.g. to decide whether it rhymes with various different words, or fits into various sentences, whether various statements were true concerning its length, its typeface, etc).

There was no standardisation of the time elapsed between the encoding task and the recognition task, not of the time allowed for the recognition task. This should not have mattered much as for any given participant it should have applied equally to each of the conditions, but if these had been standardised some random errors might have been avoided.

The participants were first year psychology students, which presented a number of problems. Firstly, all the students had studied memory theories, including the 'Levels of Processing' theory, so they would have been expecting the 'surprise' memory test. Secondly, as students they are likely to be more intelligent than average and to have better (or at last differently trained) memories. However, it is interesting that the experiment was successful even though the memory test was not unexpected. This suggests that the incidental learning paradigm is quite robust. It might be useful to confirm this by replicating the experiment but explicitly telling participants that there would be a subsequent test.

Similarly, the robustness of this paradigm could be tested further by requiring participants to say what words they could recall, rather than giving them a recognition test.

As a test of the robustness of the overall experimental findings, the experiment could be repeated without participants being told that the questions would fall into different categories. To reduce the likelihood of participants noticing this spontaneously, more conditions could be used and perhaps some questions that were simply distractors (e.g. ‘the word is one I have written in the last 24 hours’)

It should also be noted that whilst the age range was quite wide, it was not evenly spread within that range and it did not include any extreme ages. As age may affect memory, this may be a significant omission which could be investigated with a wider age range of participants and a multi-factor analysis.

The fact that words were recognised better when the answer was ‘yes’ appears to raise some difficult questions. In the case of semantic processing, at least three plausible explanations arise: Firstly, it could be argued that less processing is required when the answer is ‘no’, because the sentence could be rejected as soon as an impossible meaning is recognised, whereas when the answer is ‘yes’ it needed to be examined until the participant was sure that there was no inconsistency. Secondly, the same argument could simply imply that the word was held in consciousness for longer. Thirdly, it might be argued that when a meaningful sentence has been encountered, some trace of it remains in memory, and that this would be less likely for a non-meaningful sentence. There might be some similar effect for acoustically-processed words; in that rhymes might remain in the memory. It seems harder to explain the effect for visually-processed words; in fact for any given question it would be possible to compose one with the opposite answer which would seem to need an equal amount of processing (e.g. instead of ‘the word has less than 6 letters’, ‘the word has 6 or more letters’.) These differences become all the more perplexing when it is noted that there is no significant interaction effect between the type of processing and the yes/no condition, implying that the yes/no distinction makes the same difference in all conditions and thus a common explanation should be sought for all conditions.

This could be examined further by repeating the experiment and analysing the data for the time taken on each question, to see whether longer was taken for ‘yes’ than ‘no’ answers and whether this interacted with the type of processing. Also, between-subjects experiments could be undertaken wherein two groups had tests in which the form of the questions in the visually-processed words varied as suggested above. However, it would be necessary to check whether the different forms of the question took different times to answer.

It would also be interesting to manipulate the percentage of questions to which the correct answer was ‘yes’. It is possible that whether a word was in the majority or minority condition would influence its percentage recall. To test this sensitively, it would probably be necessary to either inform the participant that there were more answers in one condition than the other, or to disregard the early trials until he or she had noticed (or had subconsciously adjusted) to the fact.

In the rhyming and semantic conditions, participants might have started to think about possible words before one appeared. This might increased the exposure that the stimulus had (if they guessed correctly) or caused interference (if they guessed wrongly or just made several guesses). This could have been avoided by presenting the question and the stimulus at the same time.

It might have been worthwhile to analyse the words which were incorrectly judged to have been seen in the first trial – in particular, were they semantically related to, or rhymed with, words which were used in the first trial, and if this varied according to the condition (for example whether words rhyming with a target word were more likely to have been falsely recognised when the word was processed in a rhyming task). It might be possible to carry out a modification to the experiment where some words which rhymed, and were semantically related, were deliberately included in the recognition list. However it would be difficult to analyse any findings from this: it is difficult to envisage a scale whereby word A could be judged to be as semantically close to word B as word C is to word D, let alone being as close as are word E and word F which rhyme but are not semantically related.

References

Baddely, A. (1997). *Human Memory: Theory and Practice.* Hove: Psychology Press.

Eysenck, M. W. and Keane, M. T. (2000). *Cognitive Psychology: A Student's Handbook.* Hove: Psychology Press.

Glanzer, M. (1977). Storage Mechanisms in Recall. In Bower, G. (ed): *Human Memory: Basic Processes*. New York: Academic Press.

Gregg, V. (1975). *Human Memory.* London: Methuen.

APPENDICES

Appendix 1: Sample results

	Levels of processing		
	Visual	Acoustic	Semantic
Median Reaction time	1094	745	675
Error rate	0.19	0.06	0.00

	Number of words correctly recognised		
	Visual	Acoustic	Semantic
True Response	5	3	5
False response	1	4	5

Number of false recognitions = 8

Appendix 2: Pooled Results:

Participant	Number of words correctly recognised					
	True Response			False response		
	Visual	Acoustic	Semantic	Visual	Acoustic	Semantic
1	4	5	6	4	4	6
2	6	6	7	4	4	6
3	3	8	7	4	2	5
4	5	3	5	1	4	5
5	6	5	7	3	5	7
6	6	5	8	4	6	4
7	3	8	5	2	4	5
8	7	8	8	2	7	5
9	2	6	5	2	4	4
10	1	6	7	2	3	8
11	5	8	8	6	6	6
12	2	2	5	3	4	2
13	5	7	7	4	3	7
14	4	6	6	4	8	8
15	2	5	7	4	4	6
16	5	8	7	6	5	6
17	5	6	7	4	4	6
18	6	7	7	5	6	7
19	5	7	7	3	3	4
20	1	2	4	1	0	2
Mean	**4 .2**	**5 .9**	**6 .5**	**3 .4**	**4 .3**	**5 .5**
S.D.	**1 .8**	**1 .9**	**1 .1**	**1 .4**	**1 .8**	**1 .7**

Note: the Means and Standard Deviations were calculated in Microsoft Excel.

Appendix 3: Two Factor ANOVA output

Descriptive Statistics

	Mean	Std. Deviation	N
True visual	4.15	1.81	20
True accoustic	5.90	1.89	20
True semantic	6.50	1.15	20
False visual	3.40	1.43	20
False accoustic	4.30	1.78	20
False semantic	5.45	1.67	20

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

					Epsilon[a]		
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
TRUTH	1.000	.000	0	.	1.000	1.000	1.000
PROCESS	.998	.031	2	.986	.998	1.000	.500
TRUTH * PROCESS	.846	3.014	2	.219	.866	.946	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b.
Design: Intercept
Within Subjects Design: TRUTH+PROCESS+TRUTH*PROCESS

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TRUTH	Sphericity Assumed	38.533	1	38.533	24.568	.000
	Greenhouse-Geisser	38.533	1.000	38.533	24.568	.000
	Huynh-Feldt	38.533	1.000	38.533	24.568	.000
	Lower-bound	38.533	1.000	38.533	24.568	.000
Error(TRUTH)	Sphericity Assumed	29.800	19	1.568		
	Greenhouse-Geisser	29.800	19.000	1.568		
	Huynh-Feldt	29.800	19.000	1.568		
	Lower-bound	29.800	19.000	1.568		
PROCESS	Sphericity Assumed	98.150	2	49.075	39.246	.000
	Greenhouse-Geisser	98.150	1.997	49.159	39.246	.000
	Huynh-Feldt	98.150	2.000	49.075	39.246	.000
	Lower-bound	98.150	1.000	98.150	39.246	.000
Error(PROCESS)	Sphericity Assumed	47.517	38	1.250		
	Greenhouse-Geisser	47.517	37.935	1.253		
	Huynh-Feldt	47.517	38.000	1.250		
	Lower-bound	47.517	19.000	2.501		
TRUTH * PROCESS	Sphericity Assumed	3.717	2	1.858	1.071	.353
	Greenhouse-Geisser	3.717	1.733	2.145	1.071	.346
	Huynh-Feldt	3.717	1.891	1.965	1.071	.350
	Lower-bound	3.717	1.000	3.717	1.071	.314
Error(TRUTH*PROCESS)	Sphericity Assumed	65.950	38	1.736		
	Greenhouse-Geisser	65.950	32.924	2.003		
	Huynh-Feldt	65.950	35.935	1.835		
	Lower-bound	65.950	19.000	3.471		

Pairwise Comparisons

Measure: MEASURE_1

(I) PROCESS	(J) PROCESS	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1	2	-1.325*	.255	.000	-1.993	-.657
	3	-2.200*	.250	.000	-2.856	-1.544
2	1	1.325*	.255	.000	.657	1.993
	3	-.875*	.246	.006	-1.520	-.230
3	1	2.200*	.250	.000	1.544	2.856
	2	.875*	.246	.006	.230	1.520

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Appendix 4: One Way ANOVA outputs

Descriptive Statistics

	Mean	Std. Deviation	N
True visual	4.15	1.81	20
True accoustic	5.90	1.89	20
True semantic	6.50	1.15	20

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
T_CODE	.763	4.879	2	.086	.808	.872	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b.
Design: Intercept
Within Subjects Design: T_CODE

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
T_CODE	Sphericity Assumed	59.633	2	29.817	21.099	.000
	Greenhouse-Geisser	59.633	1.616	36.896	21.099	.000
	Huynh-Feldt	59.633	1.744	34.184	21.099	.000
	Lower-bound	59.633	1.000	59.633	21.099	.000
Error(T_CODE)	Sphericity Assumed	53.700	38	1.413		
	Greenhouse-Geisser	53.700	30.709	1.749		
	Huynh-Feldt	53.700	33.145	1.620		
	Lower-bound	53.700	19.000	2.826		

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	True visual - True accoustic	-1.75	2.05	.46	-2.71	-.79	-3.820	19	.001
Pair 2	True visual - True semantic	-2.35	1.42	.32	-3.02	-1.68	-7.378	19	.000
Pair 3	True accoustic - True semantic	-.60	1.50	.34	-1.30	.10	-1.788	19	.090

Descriptive Statistics

	Mean	Std. Deviation	N
False visual	3.40	1.43	20
False accoustic	4.30	1.78	20
False semantic	5.45	1.67	20

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
F_CODE	.991	.155	2	.928	.991	1.000	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b.
Design: Intercept
Within Subjects Design: F_CODE

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
F_CODE	Sphericity Assumed	42.233	2	21.117	13.426	.000
	Greenhouse-Geisser	42.233	1.983	21.298	13.426	.000
	Huynh-Feldt	42.233	2.000	21.117	13.426	.000
	Lower-bound	42.233	1.000	42.233	13.426	.002
Error(F_CODE)	Sphericity Assumed	59.767	38	1.573		
	Greenhouse-Geisser	59.767	37.676	1.586		
	Huynh-Feldt	59.767	38.000	1.573		
	Lower-bound	59.767	19.000	3.146		

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	False visual - False accoustic	-.90	1.77	.40	-1.73	-6.97E-02	-2.269	19	.035
Pair 2	False visual - False semantic	-2.05	1.70	.38	-2.85	-1.25	-5.391	19	.000
Pair 3	False accoustic - False semantic	-1.15	1.84	.41	-2.01	-.29	-2.790	19	.012

Appendix 5: Alternative Results section using One Way ANOVA results

The Means and Standard Deviations for each condition are shown in Table 2A, and the Means in Figure 1A.

Table 2A: Number of Words Correctly Recognised

	True Statement			False Statement		
	Visual	Acoustic	Semantic	Visual	Acoustic	Semantic
Mean	**4 .2**	**5 .9**	**6 .5**	**3 .4**	**4 .3**	**5 .5**
S.D.	**1 .8**	**1 .9**	**1 .1**	**1 .4**	**1 .8**	**1 .7**
% of possible words correct	52%	74%	81%	43%	54%	68%

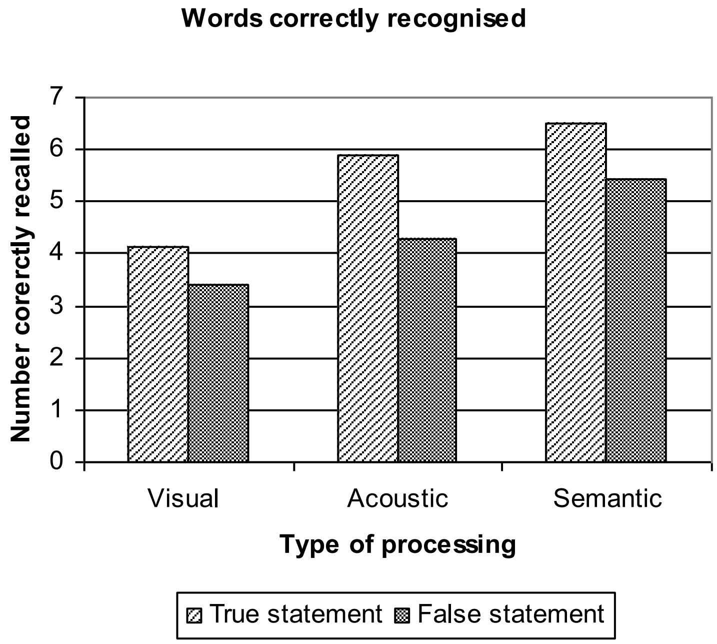

Figure 1A: Mean Number of Words Recalled

Where the answer to the original question was 'yes', a one-way repeated measures ANOVA revealed an overall significant difference between the type of processing and the mean number of words recognised [$F(2,38) = 21.099$, $p < 0.001$]. Post hoc comparisons using Bonferonni corrected paired t-tests found significant differences between visual and acoustic processing ($t = 3.820$, $df = 19$, $p = 0.001$) and between visual and semantic processing ($t = 7.378$, $df = 19$, $p < 0.001$). There was no statistically significant difference between acoustic and semantic processing.

Where the answer to the original question was 'no, a one-way repeated measures ANOVA also revealed an overall significant difference between the type of processing and the mean number of words recognised [$F(2,38) = 13.426$, $p < 0.001$]. Post hoc comparisons using Bonferonni corrected paired t-tests found significant differences between visual and semantic processing ($t = 5.391$, $df = 19$, $p < 0.001$) and between acoustic and semantic processing ($t = 2.790$, $df = 19$, $p = 0.012$). There was no statistically significant difference between visual and acoustic processing.

The hypothesis that there would be differences according to the type of processing was therefore supported. The hypothesis that there would be differences according to the answer to the processing question appears to be supported, but was not tested statistically.