The Effect of Lexicality and Word Frequency on Stroop Interference

T J Taylor
Trinity College
Cambridge

April 1992

A project report submitted for the Part II examination in Experimental Psychology

ABSTRACT

Until recently, it has been assumed that frequency effects in word naming tasks arise in the lexical identification and retrieval processes involved. However, some new experiments have suggested that the locus of such effects is subsequent to these processes. The former hypothesis predicts that frequency will affect the extent of interference produced in colour naming Stroop tasks, whereas the latter does not. In the present experiment, the size of Stroop interference using words from various frequency bands was investigated. In one condition, subjects named the colour in which high, medium and low frequency words, pronounceable non-words, or non-alphabetic character strings were displayed. In another condition, they named the printed item itself, displayed in white, for each of these item types but the last. Though substantial frequency and lexicality effects were observed in the item naming task, the amount of interference observed in the colour naming task was not significantly different for any of the real word or pronounceable non-word groups, although all of these groups did show a significant interference increment compared to colour naming of non-alphabetic character strings. It is concluded that the process of retrieval of semantic information associated with a word is not the locus of frequency sensitivity in word naming. Additionally, it is suggested that involuntary base item identification in Stroop tasks is by sublexical processes. Several problems with this view are mentioned, and the need for further research is emphasized.

INTRODUCTION

Many tasks involving word recognition, word naming and similar demands are sensitive to the frequency with which the individual word is used in the language. Such effects are well established, and many theories have been advanced concerning their origin. At present, however, experimental results have been apparently contradictory, and no consensus has been reached as to the locus of such frequency effects (see review by Monsell, 1991).

When reading a word, it is assumed that the first process that must be performed is one of identification of a 'best match' to it among all the entries in the 'mental lexicon'. When a match has been found, relevant associated semantic information is retrieved.

For *comprehension*, it is commonly assumed that two systems operate in parallel to achieve these tasks; one involving a direct link from text to meaning, and the other involving an indirect route via phonology. Phonological *activation* is considered to be relatively automatic, but phonologically-mediated *access to meaning* is slow relative to direct access, at least for high frequency words, so contributes little to the process in this case.

In tasks that require *pronunciation* of a written word, another dual-route mechanism is suspected, comprising one system which performs *word* identification and pronunciation (i.e. lexical transcoding), and another performing sublexical transcoding. In the former, the learned orthographic pattern of a known word is recognised and the associated learned pronunciation is retrieved. In parallel to this system, the sublexical system picks out recognisable pronounceable letter strings from the orthography, transcodes these using knowledge of common spelling-sound correspondences in the language, and finally assembles these into a pronunciation of the whole word.

It is typically assumed that the same set of orthographic lexical entries and the same matching process support access both from text to meaning and from text to phonology.

Large frequency effects are seen in lexical decision tasks (Rubenstein *et al.* 1970). Therefore, as it is commonly assumed that such tasks require identification but little retrieval of information, the identification stage has traditionally been seen as the major locus of frequency sensitivity.

However, some recent reports have indicated that lexical decision requires more than just identification of the word, and suggested that frequency sensitivity is due mainly to processes occurring after identification, and even after retrieval. For example, Balota and Chumbley (1984) found that a semantic categorisation task (subjects were presented with a category label, then asked whether a test word was or was not a member of the category) was much less sensitive to frequency than was lexical decision. As both these tasks require identification, Balota and Chumbley concluded that the frequency effect observed in this experiment must primarily be due to the decision process specific to the lexical decision task.

There is considerable evidence that activation of meaning also occurs during lexical decision tasks (e.g. James, 1975; Whaley, 1978; Jastrzembski, 1981; Chumbley & Balota, 1984). Balota and Chumbley's findings (*op cit.*) are therefore also consistent with the hypothesis that the process of activation of meaning is a locus of frequency effects.

Balota & Chumbley (1985) and McCann & Besner (1987) have produced evidence that in naming tasks, as in lexical decision tasks, the identification stage is not particularly frequency sensitive. Rather, they suggest that the observed frequency effects arise during the 'production' stages of the pronunciation task. McCann and Besner specify that such effects are not due to mere activation of the lexicon of phonological word forms, but rather to processes that link "lexical entries in the orthographic input lexicon with lexical entries in the phonological output lexicon."

Such a hypothesis, or, indeed, *any* hypothesis which proposes that the identification process is *not* the major locus of frequency effects is contradictory to both conventional models of lexical access and to recent connectionist models (Sejnowski and Rosenberg, 1986; Seidenberg and McClelland, 1989).

In the present experiment we have tried to study the frequency question further using a procedure based on a variation of the Stroop effect. Stroop (1935) reported that if subjects were visually presented with a colour name, and asked to name the colour of the *ink* in which the name was printed, then they were much slower if the ink colour was incongruent with the printed colour name than if it was congruent. It was later shown by Klein (1964) that colour naming of *any* common words, even if unrelated to colours, is retarded relative to naming the colour in which a row of Xs is printed, albeit to a much lesser extent than the interference originally reported by Stroop. The effect demonstrated by Klein will be referred to as the *lexical* Stroop effect throughout this report.

In Stroop's original design, blocks of words were presented on a card, and performance was measured by timing the subject at reading out the colours of all of the words. This design has subsequently been improved by presenting each word individually and measuring corresponding response times (the 'discrete trial' Stroop task). In this report, the word that is presented on each trial of a Stroop task will be referred to as the 'base word'.

The lexical Stroop effect has commonly been explained in terms of a faster (and involuntary) assignment of a spoken response to written word stimuli than to colours, and the resulting conflict between the irrelevant response to the word name and the relevant response to the colour name (Dyer, 1973; Posner, 1978). Treisman (1969) proposed that the conflict resulted from a general inability to completely focus attention on one analyzer (e.g. colour and not word-name), although this theory has been questioned by later studies (see Dyer, 1973).

Even if correct, such explanations do not tell us *why* the word-naming response is so fast compared to colour naming. In their review of Stroop experiments, Jensen & Rohwer (1966) suggest that the speed of the response arises through the large amount of practice that we have at word naming, and also through the fact that only one response is associated with each written word. A recent attempt to explain the phenomenon by Cohen *et al.* (1990), within a connectionist framework, proposes that the apparent automaticity of the word naming response arises gradually as an individual is exposed to such practice during development (the response is therefore only quantitatively, but not qualitatively, different to the colour-naming response).

The results of many experiments have suggested that, in addition to the retrieval of phonological information, Stroop interference may also be partially attributable to retrieval of associated semantic information. Warren (1972) required subjects to remember sets of words during a Stroop task. He found that interference increased when "the base word in the Stroop task was the category name for a set of words being remembered." In a later experiment (Warren, 1974) using a similar method, increased interference was observed if a word which was associatively related to the base word in the Stroop task was concurrently held in memory. Furthermore, the degree of interference was directly related to the strength of the associative connection. Other investigations have shown that Stroop tasks involve retrieval of semantic information associated with the base word by demonstrating that contextual priming affects the degree of Stroop interference observed (Conrad, 1978; Merrill *et al.*, 1981; Oden & Spira, 1983; Whitney *et al.*, 1985; Jones, 1989).

Inasmuch as the size of the lexical Stroop effect reflects the efficiency with which we (involuntarily) identify a word and retrieve associated information, it ought to be sensitive to frequency if, that is, frequency effects are primarily localized at the identification or retrieval stages.

Klein (1964) found that common words produced greater interference in a colour-naming task than did rare words. However, this finding has since been questioned because of certain methodological shortcomings in the experiment, including the fact that the rare words used were so rare that they were unlikely to have been recognized by many of the subjects. Proctor (1978) compared the results of using a response set comprising common colour words to one comprising uncommon colour words. He found the the former condition produced greater interference in a Stroop task. However, the low frequency of the uncommon words is confounded with other properties, such as the fact that they are less strongly associated with the concept of colour. This fact would, in itself, tend to reduce the amount of interference. An recent experiment by Monsell, Elliot & Perry (see bibliography) found little evidence for any frequency effect, and the significance of the result was weakened by a low accuracy for colour naming of high frequency words. Also, no control stimuli, such as XXXX, were used, and the frequency effect of the stimuli used was not demonstrated in any other task, such as word naming.

In short, no adequate experimental test of the effect of normative word frequency on interference in Stroop tasks has been conducted to date. The present report is of an experiment designed to investigate such an effect, overcoming the methodological shortcomings of previous studies.

Three categories of words were used; high frequency, medium frequency and low frequency. In order to have a base rate against which to compare colour-naming results from these categories, two control categories were used; pronounceable nonsense words and strings of non-alphabetic characters (e.g. %$£@). The latter were used rather than the more usual XXXX stimuli to attempt to match the variability between characters in this category with stimuli in the other categories, whilst maintaining roughly equal print density.

As well as giving subjects a colour naming task, an item naming task was also included (for all item categories except the non-alphabetic characters) in order to demonstrate that pronounceable stimuli in the different categories displayed the basic frequency effect in this condition.

To avoid the confounding of lexical and sublexical identification processes, only real words with irregular sound-spelling correspondences (i.e. words whose pronunciation cannot be correctly derived through knowledge of general sound-spelling patterns in English) were used in the experiment. Thus, for real-word naming, we ensured that subjects were using primarily the lexical identification route to name words, rather than the sublexical route. Naming of pronounceable non-words necessarily takes place via the sublexical route.

METHOD

DESIGN

Subjects were required to perform two tasks, item naming and colour naming (that is, naming the colour in which an item was printed), on items of four different categories; high frequency words[1] (referred to in the rest of this report as HF items), medium frequency words[2] (MF items), low frequency words[3] (LF), and pronounceable nonsense words (pseudo-words, PW). For the colour naming task, items from a fifth category, character strings (CS), were also used. The CS items comprised different combinations of the characters '@', '£', '$', '#', '&', '>', '<', '?', '*', '%' and '|'[4]. Items presented during the item naming task were in white print on a black background. Items presented during the colour naming task were in one of the colours red, green, yellow, blue or purple, again on a black background.

A discrete trial variation of the Stroop task was used to allow presentation of stimuli from the different categories in a random order.

Two separate lists of stimuli were used in the experiment, each comprising 200 items. Each list was sub-divided into groups of 40 items from the five different categories. The items used in each list are given in Appendix A.

All categories were matched for distributions of item length. Additionally, the sets of items chosen for the HF, MF, LF and PW categories were matched for number of syllables, initial sounds and stress patterns, and they all had voiced onsets to assist accurate response timing during the experiment. Across the two lists, the 80 HF, 80 MF and 80 LF items each comprised 15 verbs, 10 adjectives and 55 nouns. Equal numbers of concrete and abstract nouns were included in each list. The matching procedure is shown diagrammatically in figure 1.

Each subject was required to perform one task with each list. To avoid possible nuisance variables, subjects were counter-balanced for the order in which they were required to perform colour naming and item naming, and for which stimulus list was used for which task. There were therefore four different combinations of conditions, summarised

---

[1] High frequency words were defined as those with frequencies of 50+ per million according to the MRC database VI.3d

[2] Medium frequency words were defined as those with frequencies between 14 and 35 per million (inclusive) according to the MRC database VI.3d

[3] Low frequency words were defined as those with frequencies between 1 and 4 per million (inclusive) according to the MRC database VI.3d

[4] The '|' character actually appeared as *two* closely spaced parallel vertical lines in the experiment.

in table 1. Five subjects were sought for each condition, to give 20 subjects in total.

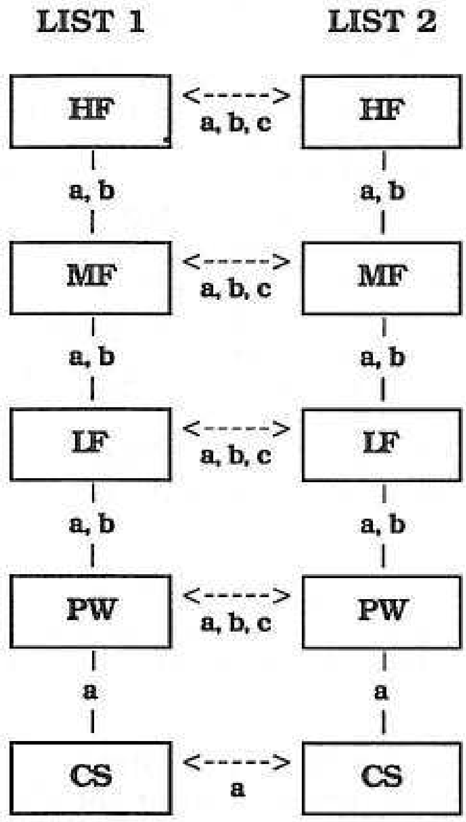

Figure 1 *The matching procedure. Where indicated, items were matched for means and distributions of **a)** length, **b)** number of syllables, and **c)** initial sounds & stress patterns*

In both tasks, items were presented in blocks of 40, so that the item naming task comprised four blocks and the colour naming task comprised five blocks.

Condition No.	First Task		Second Task	
	Response	List	Response	List
1	Item naming	1	Colour naming	2
2	Item naming	2	Colour naming	1
3	Colour naming	1	Item naming	2
4	Colour naming	2	Item naming	1

Table 1 *The four conditions used in the experiment*

For item naming, 10 items from each of the four categories were assigned in a random order to each block. For colour naming, eight items from each of the five categories were similarly assigned.

For the first subject tested in each condition (table 1), in the colour naming task each item in the source-list was assigned a colour, starting from item one assigning colours in the sequence red, green, yellow, blue, purple, red, green, etc. Thus, when the items were randomly assigned to blocks, an apparently random sequence of colours was produced. For each subsequent subject tested in each condition the initial assignment sequence of colours-to-items was shifted along by one, so that, across all five subjects in each condition, all items of the list had been presented once in each of the five colours.

SUBJECTS
Subjects were undergraduate colleagues, with ages in the range of 19 to 22 years. All subjects spoke English as their first language. No payment was offered, but subjects were told that a bottle of wine would be awarded to the person achieving the fastest times with least errors.

26 subjects were originally tested. There were problems with some subjects, the most common one being a difficulty in distinguishing between green and yellow on the monitor. In order to obtain problem-free results for five subjects in each condition, one extra subject was tested. A scatter plot of mean item naming reaction time against percentage error rate (over both tasks) was produced for the 27 subjects (see figure 2). The results were analysed using the data from 20 subjects (five from each condition) whose results fell in the shaded region of this graph. We first excluded subjects whose performance fell outside the upper time and error borders shown, and then, to make the numbers come out right, dropped one unusually fast and accurate subject and one other.

Of this final group of 20 subjects, 11 were male and 9 female.

APPARATUS
The program to run the test was written in BASIC on a BBC Model B computer. Subjects sat approximately 50cm from the screen, which was a CUB colour monitor. With this arrangement, each item in the test subtended visual angles of about 0.5° vertically and 3° horizontally. A throat microphone, connected to the computer's user port, was used to detect the subjects' speech. Detection of the onset of speech triggered the program to stop timing and hence give the subject's reaction time, to the nearest 1/100 second, for that trial.

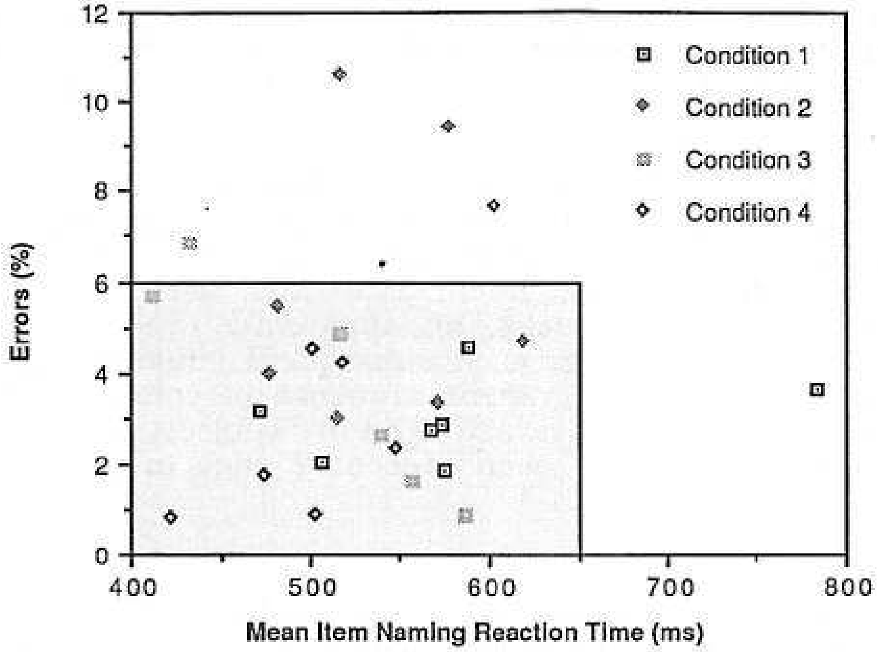

Figure 2 *Scatter plot of mean item naming reaction times against error rate for each subject*

PROCEDURE
Subjects completed both tasks in a single session, lasting 30-40 minutes.

Before each task commenced, subjects were given verbal instructions of what was required of them, and were asked to respond as quickly as possible while keeping their error rate down to 2 or 3 per 40-item block.

The subject was asked to put on the throat microphone, and its sensitivity was adjusted to ensure that the device triggered at the onset of speech but was not sensitive to swallowing, head movements etc. For the colour naming task, subjects saw words printed in each colour on the screen before the test began, so that they were familiar with the colours used.

Each task was preceded by a 15 item trial run to familiarise subjects. Additionally, each test block of 40 items contained another three extra practice items at the beginning.

On each trial, a row of white plus-signs ('+'), of the same length as the item which was about to be displayed, appeared in the centre of the screen. Subjects were asked to fixate on these. One second later, the pluses were replaced by the item in the appropriate colour. The computer measured the time from this point until the microphone detected the onset of speech in the subject. The screen was then cleared, and after a 2 second pause, the fixation symbol for the next trial appeared.

If an error was made, the experimenter signalled this to the computer by pressing a button on the keyboard. The precise type of error was recorded on a pre-printed response sheet. The data from this sheet was later added to the data file for that subject. Responses were classified into one of the following categories: correct in all respects; correct but with dubious speech onset measurement; error - fluent mispronunciation; error - wrong colour named; error - word named instead of colour; error - dysfluent; anticipation; time out (no response in 9.9 secs); or, exclude for other reasons. All data in categories other than 'correct in all respects' were ignored in the analysis of reaction times. The error rate analysis is based on all responses excluding only those correct in all respects, anticipations, timeouts and 'exclude for other reasons'.

At the end of each block, the subject's mean reaction time and number of errors were displayed. The experimenter commented encouragingly, trying to ensure that the subject always responded as quickly as possible while maintaining a low error rate. In order to minimise fatigue effects, subjects were allowed to rest between each block. However, it was found that most subjects were quite happy to continue with little interruption.

Between the two tasks, the subject was allowed to stretch his or her legs for a couple of minutes if required.

When all 20 subjects had been tested in this way, we had therefore obtained reaction times and error data for 360 items for each subject. Across all subjects, each item had appeared 10 times in each task (except CS items which were only used in the colour naming task).

RESULTS

Consistency of results over all items

In order to check whether any particular items in either of the lists gave anomalous results compared to other items in the same category, a global analysis of item naming performance and accuracy, by words, was then performed.

Of all 320 nameable items, the number of correct item naming responses obtained was below 5 out of 10 (remember, across all subjects, each item appeared 10 times in the item naming condition) for 5 items; 'AWRY', 'NARCHIE', 'OICE', 'SPOLUE' and 'HIRLCH'. Inspection of the errors that individual subjects made when naming these items revealed that 'OICE' was commonly pronounced 'DICE', and 'HIRLCH' was commonly pronounced 'HIRLICH', whereas errors in the other three items tended to be dysfluencies. Data relating to 'OICE' and 'HIRLCH' were excluded from subsequent analysis[1] because, in addition to having high error rates, these words were being *misread*, so that, at least in the case of 'OICE', they were effectively in the wrong category (e.g. a pseudoword was being read as a low frequency word). It was decided that data relating to the other three items should be retained.

Analysis of item naming latency and accuracy

The effect of item type on mean item naming reaction time was reliable at the $p<.0001$ level both by subjects and by items ($F_{subjects}(3,48)=78.4$, $F_{items}(3,312)=47.7$). Pairwise differences in item naming reaction time between each of the item types were all significant at the .01 level (Newman-Keuls test). The mean reaction times for each item type are shown in figure 3 (left-hand side).

There were no significant main effects for order (i.e. whether item naming was the first or second task) or list (i.e. whether the item belonged to list 1 or 2).

This analysis was repeated using median, rather than mean, reaction times, and the results followed the same pattern, suggesting that the data was not affected by outliers.

The anova table for the by-subject analysis is shown in Appendix B.

---

[1] This was the theory. After completion of the analysis of this experiment, it came to light that, due to inconsistencies in item numbering between two types of data file used in the analysis, data relating to the item 'LOAK' was discarded instead of that relating to 'HIRLCH'. Time did not permit a complete reanalysis prior to the deadline for submission of this report. However, as this represents only one item of the 40 from the pseudoword category of list 2, and one of the 80 pseudowords in total, and additionally as 'OICE' was the only item that was being misread as a *real* word, it was considered that this error would have negligible effect on the results.

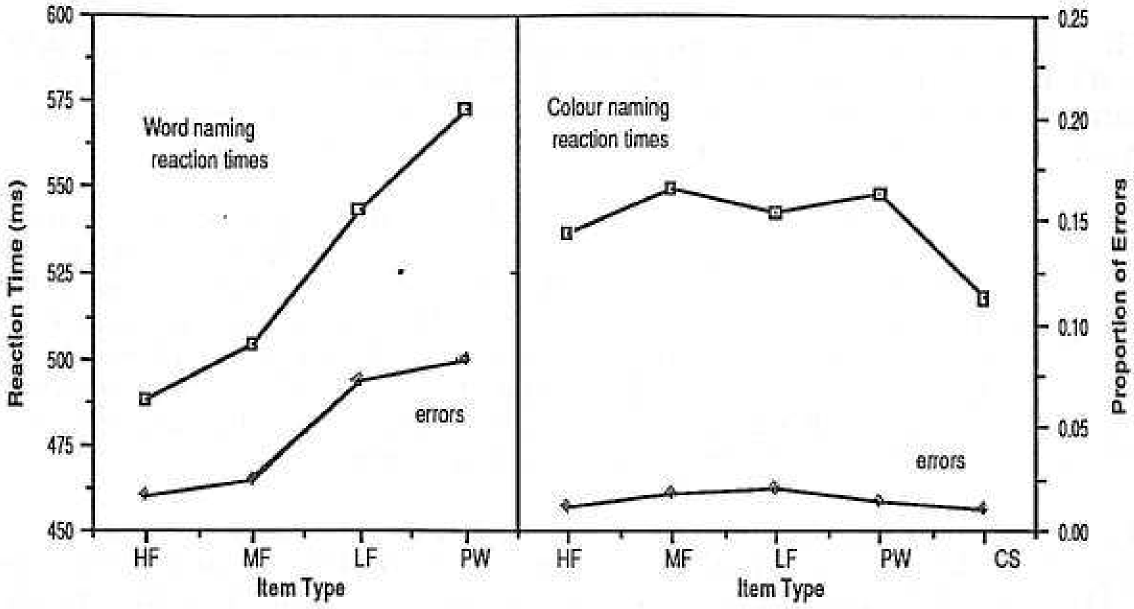

Figure 3 *Mean correct item naming (left) and colour naming (right) reaction times and associated proportion of errors*

The effect of item type on proportion of errors in the item naming task was also reliable at the $p<.0001$ level both by subjects and by items ($F_{subjects}(3,48)=19.5$, $F_{items}(3,312)=8.4$). The mean proportions of errors are plotted in figure 3 (left-hand side).

The difference between mean proportion of errors for MF and LF items was significant at the .01 level (Newman-Keuls test), whereas the differences between HF and MF, and between LF and PW were insignificant.

An interaction effect between list and type also approached significance ($F(3,38)=2.8$, $p=.0513$) in the by-subject analysis, but not in the by-item analysis.

The anovas on the error data were performed on an arcsine transformation of proportion of errors in order to create a more normal distribution of data.

Analysis of colour naming latency and accuracy

As for item naming, the effect of item type on colour naming was reliable at the $p<.0001$ level both by subjects and by items ($F_{subjects}(4,64)=19.1$, $F_{items}(4,388)=15.3$). The mean reaction times for each item type are shown in figure 3 (right-hand side). The difference between mean reaction time for CS items was significantly different to means for all other items at the .01 level (Newman-Keuls test). No other inter-item differences were significant at this level.

The effect of colour on colour naming was also reliable at the p<.0001 level ($F_{subjects}(4,64)=35.4$). All differences between mean reaction times for individual colours were significant at the .01 level (Newman-Keuls test) except the difference between yellow and blue.

The interaction between item type and colour in the colour naming task was also significant ($F(16,256)=2.5$, $p=.0014$) - see figure 4. There was considerably more variation in mean reaction times of the different item types for some colours (notably, green and purple) than for others. However, looking at individual colours, *there is no significant or consistent effect of lexicality (i.e. PW items compared with HF, MF and LF items) or word frequency (i.e. difference between HF, MF, LF items) on mean colour naming reaction time.*

The effect of task order on colour naming was not significant. However, list was a significant main effect in the by-items analysis ($F(1,388)=16.6$, $p=.0001$), although it was insignificant in the by-subjects analysis. For all item types, colour naming mean reaction times were consistently about 5-20ms slower for list 1 items than for list 2 items. This could have come about if several of the subjects assigned list 1 for colour naming were particularly slow. If larger numbers of subjects were tested, we would expect this difference to disappear. No consistent or significant difference was found in the colour naming mean reaction time of each colour between the lists.

The anova table for the by-subject analysis is shown in Appendix B.

The analyses were repeated using median reaction times, and similar results were obtained.

No significant main effects or interactions were found for the proportions of errors in the colour naming task. (The anova was performed on an arcsine transformation of the corresponding proportions of errors to provide data with a roughly normal distribution). The mean proportion of errors for each item type is shown in figure 3 (right-hand side).

Analysis of overall reaction times, by subject

The data was analysed to test for performance differences between the tasks of item naming and colour naming.

The effect of task on mean reaction times was highly significant ($F(3,48)=64.2$, $p<.0001$). There were no other significant main effects between the two tasks.

A significant interaction was found between task and item type ($F(3,48)=44.3$, $p<.0001$). This interaction can be seen in figure 3, but is depicted in a clearer fashion in figure 5.

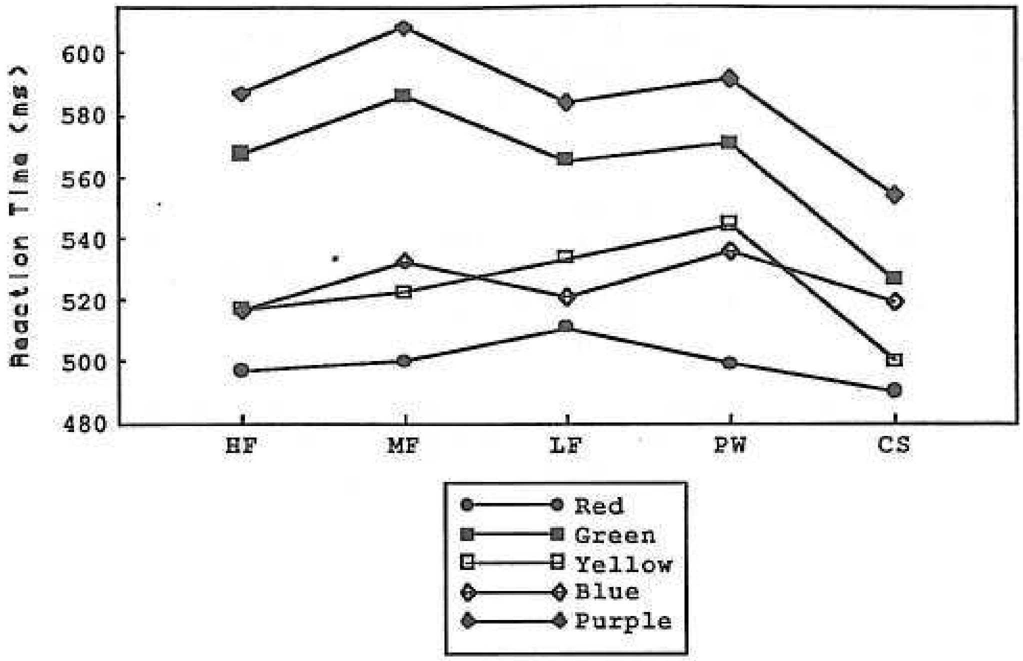

Figure 4 *Interaction between Colour and Type for mean correct colour naming reaction times*

The familiar reverse proportionality between word frequency and colour naming time is clearly demonstrated.

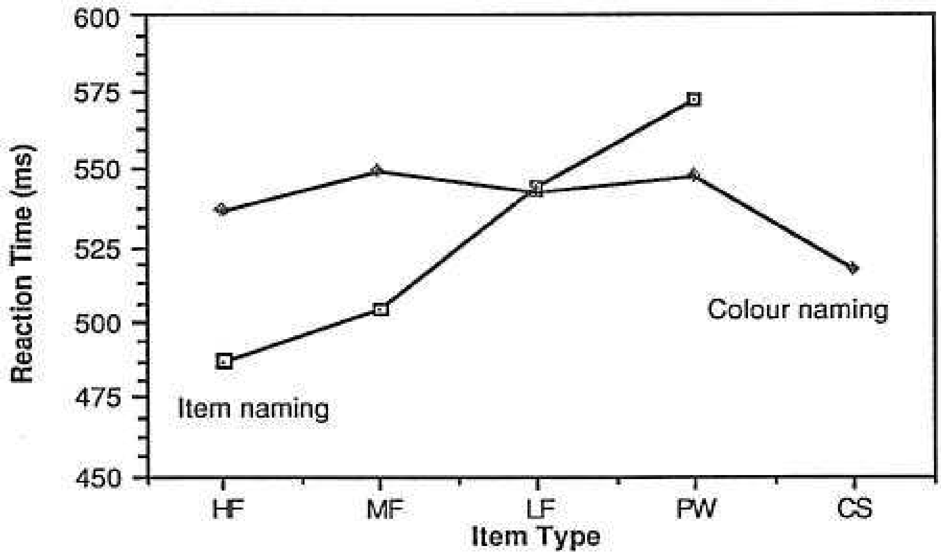

Figure 5 *Interaction effect between item type and task for mean correct reaction times*

However, as found in the previous section, there is no statistically significant difference in colour naming mean reaction times between HF, MF, LF and PW item types, although these are all significantly slower than colour naming mean reaction times for CS items.

It is also notable that for LF and PW items, colour naming was actually faster (albeit only marginally in the case of LF items) than item naming. The difference in mean correct reaction times between item naming and colour naming is significant at the .01 level for all item types except LF (Newman-Keuls test).

By-subject anovas on mean item naming reaction times and mean colour naming reaction times which included the sex of the subjects as a factor revealed that this was not a significant main effect.

DISCUSSION

The main finding of the experiment is that there are no significant differences between colour naming reaction times for real words and pronounceable non-words, i.e. no lexicality effect was observed. Given this, it is not surprising that we also found no significant differences in the amount of Stroop interference between high-, medium- and low-frequency words, i.e. no frequency effect was observed.

The same words and non-words were shown to produce significant frequency and lexicality effects between groups for item-naming reaction times, in the expected manner.

These results suggest that either frequency effects arise subsequent to the identification and lexical access processes involved in the involuntary processing of the word in a Stroop task, or that Stroop interference is due to something other than naming response competition.

The former hypothesis is not without support, as mentioned in the introduction and reviewed by Monsell (1991). However, the present results suggest that the process of retrieval of semantic information relating to the base word, that has been shown to occur in some Stroop tasks, is not frequency sensitive. As a strong frequency effect was found in the item naming condition, which presumably involves processes of identification and retrieval of semantic and phonological information, then these results are consistent with McCann and Besner's (1987) proposal that a major locus of frequency sensitivity is within the production stages of the item-naming task.

If these assumptions are correct, then the results represent a serious challenge both for conventional models and for recent connectionist models of lexical access, which commonly attribute frequency sensitivity largely to the lexical identification process.

Equal interference was observed in colour naming of all categories of words and pronounceable non-words, and this was significantly greater than the interference in colour naming of non-alphabetic control items. It is therefore tempting to conclude that a sublexical route to the assignment of the involuntary item naming response is being utilised in these Stroop tasks. Such a route is, as already stated, insensitive to word frequency. This conclusion is problematic when one remembers that the real words used in the experiment all had irregular sound-spelling correspondences, so that a sublexical process could not achieve a definite identification of these words. However, one could imagine a process whereby sublexical transcoding leads to activation of several entries in the phonological lexicon, and the first entry to achieve a certain threshold activation is selected for pronunciation. This hypothesis does not stipulate that the *correct* lexical entry is selected. It would be of interest to perform an experiment involving priming of base words in a Stroop task to see whether a condition that will prime a word with similar phonological properties to the base word (but not associated to it semantically) can affect the amount of Stroop interference observed.

The fact that item naming of pronounceable non-words was quicker than colour naming for the same items also represents a potential problem for this account, as the same process of sublexical transcoding is assumed to be operating in both cases. It could be supposed that, in this case, the colour-naming response is available before the item naming response. (It is assumed on the basis of experiments involving priming of the base word by semantically associated words, that for colour naming of real words, the item identification and naming response is available before the colour naming response). If this is so, then it would be expected that if subjects were tested for subsequent recognition of items used in the colour-naming task, their performance would be substantially better for real words than for pronounceable non-words.

The possibility remains that the present results reflect the fact that Stroop interference is not due to naming response competition at all. The fact that colour naming of pronounceable non-words is significantly slower than colour naming of non-alphabetic character strings implies that interference due to processing of the item is occurring. However, it has just been suggested that for colour naming of pronounceable non-words, the colour naming response is available before the item naming response. If this is true, then any interference due to processing of the item must arise in a stage prior to production of the response. Such an interpretation appears contradictory to studies such as that of Keele (1973), which have shown that words that produce a small interference over non-word controls when a vocal response is required for the colour name, produce no such interference when a key press output is used. McCann and Besner's suggestion (1987) that the primary locus of frequency effects is in processes linking lexical entries in the orthographic input lexicon with entries in the phonological output lexicon could resolve this apparent contradiction, if one assumes that a corresponding *linking* process for sublexical units *had* occurred in colour naming of pronounceable non-words before the colour naming response became available. As this process involves linking of sublexical portions of an item from orthography to phonology, this would also explain why no *word* frequency effect was observed for real words.

In conclusion, our results suggest that the process of retrieval of semantic information associated with a word is not the locus of frequency sensitivity in word naming. Additionally, they imply the possibility that involuntary base item identification in Stroop tasks is by sublexical processes. However, several problems with this view have been mentioned. Further research, including the suggestions made in this section, is required to give us a better understanding both of loci of frequency sensitivity and of the precise processes that lead to Stroop colour naming interference.

APPENDIX A: ITEMS USED IN THE EXPERIMENT

HIGH FREQUENCY		MEDIUM FREQUENCY	
LIST 1	LIST 2	LIST 1	LIST 2
ACCOUNT	AMOUNT	ANCHOR	ADJUST
ADDRESS	ATTEMPT	ATTRACT	ALERT
ADVICE	ANSWER	ALARM	ARGUE
ARE	ANCIENT	AUNT	ANGEL
BEAR	BEAUTY	BOWL	BUSH
BLOOD	BUILD	BOLD	BEHALF
BOTH	DEATH	BULL	DECAY
DISEASE	DOOR	DIET	DELIGHT
DOUBT	DIRECT	DIVINE	DEFEND
GROSS	COLOUR	GUILT	COURAGE
COUNTRY	CONCERN	CLERK	CHORUS
COUPLE	COME	CAUTION	CHAOS
CREATE	COVER	COLLECT	CRUEL
EYE	EXIST	EXACT	EXTEND
EXPECT	EXTREME	EXCEED	ENGAGE
HEAD	HOTEL	HONEY	HEIGHT
HEAVY	HERE	HATRED	HEALTHY
HONOUR	IDEA	OCEAN	ONION
INCLUDE	LEAGUE	INTEND	LOVER
LEARN	MIND	LAUGH	MONK
MONEY	MONTH	MILEAGE	MOTIVE
MOTHER	MANY	MUSEUM	MILD
NOTHING	PEOPLE	NUDE	PATROL
POST	PRETTY	PRAYER	PARADE
PROVE	POLICE	PURSUIT	PRECISE
PROCESS	CIRCLE	PROMOTE	CEASE
SAMPLE	SOURCE	SWEAT	SEWAGE
SUCCESS	SOCIAL	SECURE	SUBTLE
SUPREME	SPREAD	SELECT	SUE
SUGGEST	SUPPORT	SERUM	SCHOLAR
FIGURE	FOOT	FOLK	FLOOD
FRIEND	FRONT	FINANCE	FAULT
FATHER	TRUTH	FRUIT	TONGUE
TROUBLE	VERY	TOURIST	VARY
VARIOUS	WATER	VAGUE	WARRANT
WATCH	WOMAN	WEALTH	WOUND
WORK	WORRY	WHISKEY	WARD
WANT	REGION	WORST	REALM
RESULT	REMAIN	ROUTINE	RHYTHM
REMAIN	REMOVE	ROLL	RESIST

LOW FREQUENCY		PSEUDOWORDS	
LIST 1	**LIST 2**	**LIST 1**	**LIST 2**
ABODE	ABYSS	ALLUM	AMSUTT
AGUE	ASTHMA	ARINT	ARPLE
ATTEST	AWRY	AGGISH	AMTRY
ARCTIC	ACHE	APPENT	ANE
BLOUSE	BOUGH	BENNEL	BERKIN
BEIGE	BEGUILE	BISTLE	BLEAN
BINDER	DEARTH	BRAIF	DAPER
DEMISE	DRAUGHT	DELTH	DESTO
DUNGEON	DEFUNCT	DISSION	DANDEN
GRANITE	CADET	GRELKER	CRONTH
CANAL	CASTE	CREASH	COPLE
CHASM	COUPON	CLEEDLE	CLABBY
KHAKI	COERCE	CROMP	KROWN
ERR	EXTINCT	EATTOCK	ENUE
EXCITE	EXALT	ERNELL	ENTLE
HAREM	HEARSE	HENSATE	HOID
HEARTY	HOMAGE	HULF	HIRLCH
ORDEAL	IDLER	OLPITH	OICE
INFER	LACQUER	ISSER	LOAK
LAMENT	MALICE	LURNEY	MARDLER
MISHAP	MALIGN	MENTY	MISP
MONARCH	MONSOON	MOBE	MUNKER
NAUSEA	PALL	NARCHIE	PABE
PALSY	PIGEON	PARMOIL	PETTIC
PROPEL	PROFUSE	PLONNET	PORTISH
PERPLEX	SIEVE	PRAIK	SOIT
SLUICE	SURFEIT	SLAMP	SLUBE
SWAN	SWARM	SNEAPER	SAVVER
SUBDUE	SUPINE	SWITTON	SALLEN
PSALM	SEDATE	SLITH	SPOLUE
FAMINE	FEUD	FIPPICK	FONDULE
FLORIST	FOREARM	FOODLE	FRAITH
FETE	TROUGH	FREAM	TIRY
TIGRESS	VASE	TRINTER	VOAD
VOW	WAND	VOLLER	WEF
WOMB	WORM	WOSSLE	WENDO
WARP	WAN	WOZE	WURDER
WHARF	RECEIPT	WILNET	RELK
RENOWN	REPUTE	RHISTLE	RHORTLE
ROUSE	ROGUE	RIDDON	RIMBLE

CHARACTER STRINGS

LIST 1

#&*£$??
&&$>%
>&@&£&|
&|*?|#&
?><>&%&
|@$*%>
|><£%&|
@£|>@@>
>*<*@?
|?<<>@
&£|<*<
@£#&%|
%%*%#<
@>&£%>
>?|>?@
@|&@$*
*$%@|£
#%<£**
|*%%&&
<%£?
|%*<><
|£@&#
>&*@|
<@>&%
<&@%£
?@??%
|£%#&
#*<|?
<>*%?
&>$$|
||>*@
*|&@<
@?<?
$$##
$&#>
<%@>
@@%<
#|#%
*@<%
£*>@

LIST 2

#&*£$??
&&$>%
>&@&£&|
&|*?|#&
?><>&%&
|@$*%>
|><£%&|
@£|>@@>
>*<*@?
|?<<>@
&£|<*<
@£#&%|
%%*%#<
@>&£%>
>?|>?@
@|&@$*
*$%@|£
#%<£**
|*%%&&
<%£?
|%*<><
|£@&#
>&*@|
<@>&%
<&@%£
?@??%
|£%#&
#*<|?
<>*%?
&>$$|
||>*@
*|&@<
@?<?
$$##
$&#>
<%@>
@@%<
#|#%
*@<%
£*>@

Note:

1) As character strings were only used for the colour-naming task, the same 40 items were used in both lists.
2) The '|' character used in this report actually appeared on the screen as *two* closely spaced parallel vertical lines.

APPENDIX B: SELECTED ANOVA TABLES

Source of variation	df	Sum of squares	Mean square	F	p	Epsilon correction
ANOVA on Mean Correct Item Naming Reaction Times, by Subject						
O	1	9090.848	9090.848	.803	.3834	
L	1	237.361	237.361	.021	.8867	
OL	1	2400.241	2400.241	.212	.6514	
Error	16	181102.239	11318.890			
T	3	88690.475	29563.492	78.372	**.0000**	
OT	3	1008.255	336.085	.891	.4526	
LT	3	253.323	84.441	.224	.8794	
OLT	3	42.337	14.112	.037	.9902	
Error	48	18106.521	377.219			.51
ANOVA on Mean Correct Colour Naming Reaction Times, by Subject						
O	1	2159.042	2159.042	.031	.8629	
L	1	18141.878	18141.878	.259	.6180	
OL	1	139264.698	139264.698	1.985	.1780	
Error	16	1122461.874	70153.867			
T	4	67265.531	16816.383	19.112	**.0000**	
OT	4	3955.303	988.826	1.124	.3532	
LT	4	1671.950	417.987	.475	.7539	
OLT	4	521.268	130.317	.148	.9632	
Error	64	56313.657	879.901			.76
C	4	468572.007	117143.002	35.413	**.0000**	
OC	4	20724.139	5181.035	1.566	.1941	
LC	4	7904.211	1976.053	.597	.6658	
OLC	4	13151.834	3287.959	.994	.4174	
Error	64	211707.186	3307.925			.63
TC	16	37099.890	2318.743	2.510	**.0014**	
OTC	16	13380.944	836.309	.905	.5638	
LTC	16	7370.329	460.646	.499	.9468	
OLTC	16	15416.081	963.505	1.043	.4118	
Error	256	236517.327	923.896			.45

Significant effects are shown in **bold**

BIBLIOGRAPHY

Balota, D. A., Chumbley, J. I., "Are Lexical Decisions a Good Measure of Lexical Access? The Role of Word Frequency in the Neglected Decision Stage" *Journal of Experimental Psychology: Human Perception and Performance* 1984;10:340-357

Balota, D. A., Chumbley, J. I., "The Locus of Word-Frequency Effects in the Pronunciation Task: Lexical Access and/or Production?" *Journal of Memory and Language* 1985;24:89-106

Chumbley, J. I., Balota, D. A., "A Word's Meaning Affects the Decision in Lexical Decision" *Memory and Cognition* 1984;12:590-606

Cohen, J. D., Dunbar, K., McClelland, J. L., "On the Control of Automatic Processes: A Parallel Distributed Account of the Stroop Effect" *Psychological Review* 1990;97:332-361

Conrad, C., "Some Factors Involved in the Recognition of Words" *in* J. W. Cotton & R. L. Klatzky (eds.) "Semantic Factors in Cognition" 1978, New York: John Wiley & Sons

Dyer, F. H., "The Stroop Phenomenon and its Use in the Study of Perceptual, Cognitive and Response Processes" Memory and Cognition 1973;1:106-120

James, C. T., "The Role of Semantic Information in Lexical Decisions" *Journal of Experimental Psychology: Human Perception and Performance* 1975;1:130-136

Jastrzembski, J. E., "Multiple Meanings, Number of Related Meanings, Frequency of Occurrence, and the Lexicon" *Cognitive Psychology* 1981;13:278-305

Jensen, A. R., Rohwer, W. D. Jr., "The Stroop Color-Word Test: A Review" *Acta Psychologica* 1966;25:36-93

Jones, J. L., "Multiple Access of Homonym Meanings: An Artifact of Backward Priming?" *Journal of Psycholinguistic Research* 1989;18:417-432

Keele, S. W., "Attention and Human Performance" 1973, Pacific Palisades, California: Goodyear

Klein, G. S., "Semantic Power Measured Through the Interference of Words with Color-Naming" *American Journal of Psychology* 1964;77:576-588

McCann, R. S., Besner, D. "Reading Pseudohomophones: Implications for Models of Pronunciation Assembly and the Locus of Word Frequency Effects in Naming" *Journal of Experimental Psychology: Human Perception and Performance* 1987;13:13-24

Merrill, E. C., Sperber, R. D., McCauley, C., "Differences in Semantic Encoding as a Function of Reading Comprehension Skill" *Memory and Cognition* 1981;9:618-624

Monsell, S., "The Nature and Locus of Word Frequency Effects in Reading" *in* D. Besner & G. W. Humphreys (eds.) "*Basic Processes in Reading: Visual Word Recognition*" 1991, Hillsdale N.J.: Lawrence Erblaum Associates

Monsell, S., Elliot, R., Perry, N., "The 'Emotional' Stroop Effect" *unpublished Part II report* Department of Experimental Psychology, Cambridge

Oden, G. C., Spira, J. L., "Influence of Context on the Activation and Selection of Ambiguous Word Senses" *Quarterly Journal of Experimental Psychology* 1983;35A:51-64

Proctor, R. W., "Sources of Color-Word Interference in the Stroop Color-Naming Task" *Perception & Psychophysics* 1978;23:413-419

Posner, M. I., "Chronometric Explorations of Mind" 1978, New York: Oxford University Press. pp 99 ff

Rubenstein, H., Garfield, L., Millikan, J. A., "Homographic Entries in the Internal Lexicon" *Journal of Verbal Learning and Verbal Behavior* 1970;9:487-494

Seidenberg, M. S., McClelland, J. L., "A Distributed, Developmental Model of Word Recognition and Naming" *Psychological Review 1989*

Sejnowski, T. J., Rosenberg, C. R., "NETtalk: A Parallel Network that Learns to Read Aloud" Report 13 of the Cognitive Neuropsychology Laboratory, The John Hopkins University

Stroop, J. R., "Studies of Interference in Serial Verbal Reactions" *Journal of Experimental Psychology* 1935;18:643-661

Treisman, A. M., "Strategies and Models of Selective Attention" *Psychological Review* 1969;76:282-299

Warren, R. E., "Stimulus Encoding and Memory" *Journal of Experimental Psychology* 1972;94:90-100

Warren, R. E., "Association, Directionality and Stimulus Encoding" *Journal of Experimental Psychology* 1974;102:151-158

Whaley, C. P., "Word-Nonword Classification Time" *Journal of Verbal Learning and Verbal Behavior* 1978;17:143-154

Whitney, P. W., McKay, T., Kellas, G., Emerson, W. A. Jr., "Semantic Activation of Noun Concepts in Context" *Journal of Experimental Psychology: Learning, Memory and Cognition* 1985;11:126-135