

Naming the color of a word: Is it responses or task sets that compete?

STEPHEN MONSELL, TIM J. TAYLOR, and KAREN MURPHY
University of Cambridge, Cambridge, England

Subjects named the colors in which high- and low-frequency words and pronounceable nonwords, otherwise matched, were displayed. Color naming was slower for all three item types than for visually equivalent strings of nonalphanumeric symbols but was no slower for words than for nonwords, nor for high-frequency words than for low-frequency words. Unpronounceable letter strings had intermediate color-naming latencies. However, frequency and lexical status had large effects on latency for reading the same words and pseudowords aloud. Interference is thus predicted not by the strength of association between a letter string and its pronunciation but by the presence of word-like constituents. We argue that the interference from an unprimed noncolor word is due to, and isolates, one of two components of the classic Stroop effect: competition from the whole task set of reading. The other component, response competition, occurs only when lexical access is sufficiently primed.

As every psychology student is taught, it is harder to name the color in which a letter string is displayed if that letter string happens to be the name of a conflicting color word (e.g., GREEN displayed in red) than if it is a control string (e.g., XXXXX displayed in red). This effect, reported by and named after Stroop (1935), has generated an enormous research literature (see MacLeod, 1991, for review). In a recent article, MacLeod (1998) asked, rhetorically, “Why is this phenomenon so compelling?” and answered “This task provides a theoretical window on how we deal with conflicting stimuli and task demands, and is a fertile testing ground for ideas about automaticity and the role of learning in the development of that automaticity. These are fundamental questions about how attention works” (p. 201). We agree. The question we address in this article is whether the interference results from competition between specific responses associated with the stimulus or from conflict between the task sets evoked by instructions, context, and stimuli.

Interference From Noncolor Words

Klein (1964) reported that even a noncolor word causes interference with color naming. Figure 1 shows data from Klein and a replication by Fox, Schor, and Steinman (1971). Both studies measured the time taken to name the color of each of 80 items printed on a card in a random

mixture of four colors. There were different cards for different conditions. For the baseline condition, the items were strings of asterisks (Klein, 1964) or Os (Fox et al., 1971). In both studies, interference was greatest in the standard Stroop condition, in which the items were the names of the colors in the response set. A declining gradient of interference was seen across the other conditions, in the following order: color words not in the response set, words associated with a response set color (e.g., *lemon*, *grass*), common and rare words with no color associations, and consonant strings. We focus here on the interference observed for words with no color associations, which we refer to as the *Klein effect*. This name may also remind us that the effect is usually smaller than the classic Stroop effect—although MacLeod (1996) reports a discrete-trial experiment in which unrelated noncolor words, presented once only, were color-named 119 msec slower than XXXX controls.

Competition Between S–R Associations?

Interference with color naming is usually interpreted as due to lexical access activating a potential response competing with the production of the color name (e.g., Dyer, 1973; Posner & Snyder, 1975). Skilled readers have available a well-practiced (some would say “automatic”) chain of processes linking visual fixation on a familiar word to retrieval of its meaning and pronunciation (the left-hand pathway in Figure 2.) Color naming requires us to activate a color name from the same input but via a different pathway—identification of the color and retrieval of its name, directly and/or via semantics (the right-hand pathway of Figure 2). The strong interference observed for color words—the classic Stroop effect—presumably occurs because both the desired response and the competitor are members of the same category, making it especially difficult to select the appropriate response. This

Experiment 1 was conducted as a third-year undergraduate research project by T.J.T. and K.M. under the supervision of S.M. Experiments 2 and 3 were run by T.J.T. with the aid of a Grindley Grant from the University of Cambridge. The authors thank Colin MacLeod and an anonymous reviewer for their comments on an earlier draft and Nick Yeung for much helpful discussion of the relation between task set and response competition. Correspondence should be addressed to S. Monsell, who is now at the School of Psychology, University of Exeter, Exeter EX4 4QG, U.K. (e-mail: s.monsell@exeter.ac.uk).

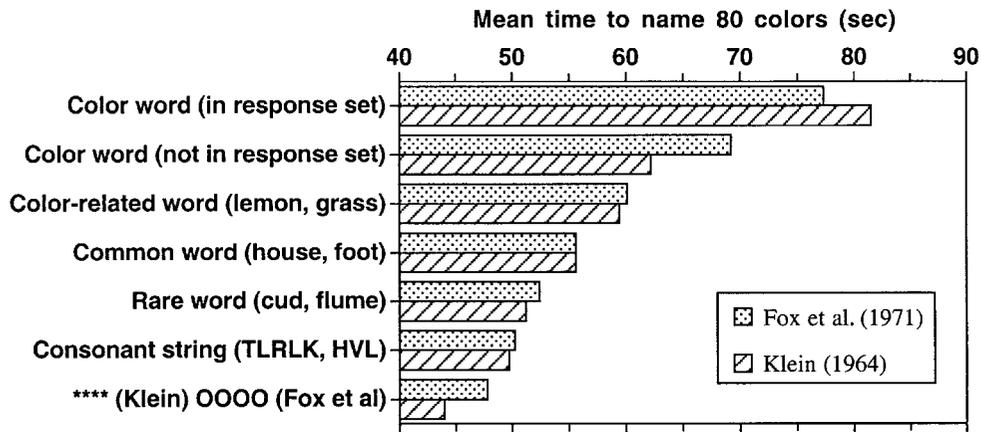


Figure 1. Mean time (in seconds) to name the colors of 80 items of different types in the experiments of Klein (1964) and Fox et al. (1971).

extra difficulty is ameliorated, but not eliminated, for a color word never used as a response. Spread of activation to color concepts associated with a word is assumed to cause greater activation of competing color names than for words with no color associations. Hence, the gradient of interference effects observed by Klein (1964) and Fox et al. (1971) can be understood in terms of a gradient of activation of the inappropriate response.

The attribution of the interference to competition between individual responses associated with the stimulus may be illustrated by Cohen, Dunbar, and McClelland's (1990) connectionist model of the Stroop effect (see also Cohen, Usher, & McClelland, 1998; Kanne, Balota, Spieler, & Faust, 1998). The model has unidirectional connections from "ink color" and "word" units via separate sets of hidden units to "response" units (one for each word in the response set). Each set of hidden units also receives input from a corresponding task unit. Activation of one or the other task unit determines whether the response units will be activated primarily by input from color or from word units. However, this biasing input interacts with the strength of association between input and output units. A lifetime of reading makes the word-to-response associations much stronger than the color-to-response associations. Although activation of the "word task" unit biases the hidden units sufficiently to suppress activation of the response units by ink color units, input to the hidden units from the "color task" unit cannot prevent some activation of response units by the word units. This explains the asymmetry of interference: Reading a color word aloud usually shows little or no interference from a conflicting color. A major motivation for the model was a demonstration (MacLeod & Dunbar, 1988) that when participants were taught arbitrary associations between shapes and color names, incongruent color initially interfered with shape naming, whereas incongruent shape had no effect on color naming; after 20 days of practice at labeling shapes with color names, this asymmetry of interference

reversed (see also Dulaney & Rogers, 1994). Stroop-like interference thus appears to be a continuous function of the relative strength, as modulated by practice, of competing stimulus-response (S-R) associations. An alternative explanation—that practice merely changes the relative speed of processing in the two pathways—has generally been discounted following a demonstration (Glaser & Glaser, 1982) that presenting the color substantially in advance of the word did not reverse the asymmetry of interference.

Cohen et al.'s (1990) model exemplifies a theory in which interference is predicted by the relative strengths of individual S-R associations: those from the stimulus to task-relevant and task-irrelevant responses. The stronger the task-relevant association, the greater the interference. Applying this principle to the Klein effect, two straightforward predictions follow, given that associative strength is a function of experience. (1) A familiar word clearly has stronger associations between its form and its name (both directly and via meaning) than does a pronounceable nonword. It should therefore take longer to name the color of the word, all other things being equal. (2) A high-frequency word should cause more interference (a larger Klein effect) than a low-frequency word.

The latter prediction depends on the locus of frequency effects in word naming. Most theorists assume that frequency influences the time taken to activate a phonological or semantic code associated with the orthographic form (Monsell, 1990, 1991; Monsell, Doyle, & Haggard, 1989; Plourde & Besner, 1997), because frequency influences the resting level of activation of word-form detectors (Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981) or the strength of connections (Borowsky & Besner, 1993; Plaut, McClelland, Seidenberg, & Patterson, 1996) or a verification-checking process (Paap & Johansen, 1994). If, instead, frequency effects on naming were located (at least in part) at a later stage of articulatory production (e.g., Balota & Chumbley, 1985; see also Goldinger,

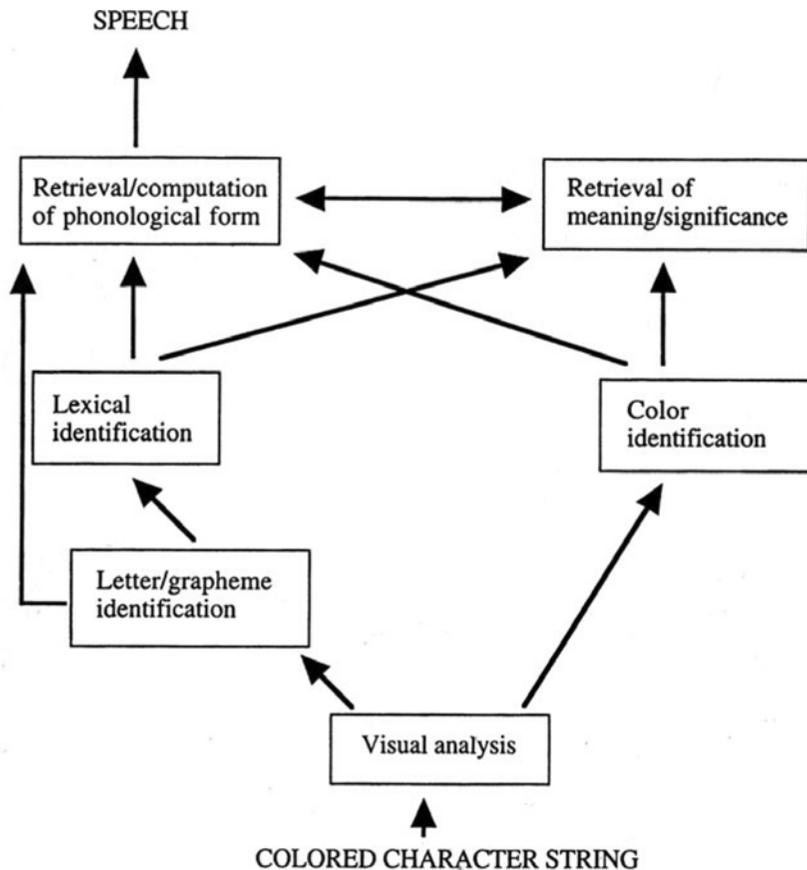


Figure 2. Pathways between sensory encoding of a word and activation of a pronunciation.

Azuma, Abramson, & Jain, 1997), then the competition between response codes responsible for the Klein effect might escape (some of) the effects of frequency. But, even then, we would still expect a substantial difference in the interference caused by words and nonwords: Only the former possess learned pronunciations.

Competition Between Task Sets?

We now consider the possibility that interference with color naming might result from competition not only between individual response codes associated with the stimulus but also from competition between task sets.

As a visible object with a familiar form, a name, and a meaning, a written word affords many possible tasks: naming, lexical decision, semantic classification, translation, free association, and color naming, among others. Before a word appears, one can set oneself to perform any of these tasks, activating (for a familiar task) or constructing (for an unfamiliar one) the appropriate configuration of cognitive processes—a “task set.” For naming, this will require activating and enabling the transmission of information along the left-hand pathway depicted in Figure 2 and setting the parameters of its processes so that the word will be named rather than associated to or translated, in a

normal voice not a shout, and as quickly as is compatible with an acceptable degree of accuracy and articulatory fluency. We carry many learned task sets in our heads. They can be evoked with just a brief instruction, such as “Please name each digit,” but they vary in their degree of practice and automaticity. Task sets for unfamiliar tasks can be installed by more elaborate verbal instructions (as in many psychology experiments) and/or by trial-and-error learning. In associationist language, a task set is a whole set of S–R associations. In the language of symbolic computation, an active task set is a set of instructions or condition–action rules held in procedural working memory (see Meyer & Kieras, 1997, for explicit modeling of examples).

The notion of “task set” is essential in explaining why we do not always name (or translate, or classify) attentively fixated words. Little is known about how task sets are acquired, represented, and reconfigured, but there is a line of research based on the observation that when one switches between two tasks, reaction time (RT) on the first trial after a switch is prolonged. This switch cost is attributed to a process, or consequences of a process, of reconfiguring task set. Two ideas from this literature will be helpful here. One is that whole task sets compete, over and

above any competition between specific responses associated with a stimulus. The other is that the activation of task sets is driven, in part, by stimuli and by endogenous voluntary control.

One source of evidence for competition between task sets is the positive and negative priming of task sets labeled *task-set inertia* (TSI) by Allport, Styles, and Hsieh (1994). They showed, and Allport and Wylie (2000) confirmed, that performance of a task is harder if another task afforded by the stimuli has been performed in the last few minutes. They argue that, once a task has been performed, activation of its task set and inhibition of competing task sets carry over to subsequent trials and interfere with response selection, even when the stimuli and responses are different. As well as demonstrating a relatively long term TSI effect, Allport and colleagues have also argued that a more transient component of TSI accounts for the “residual” cost of task switching—the longer RT observed on the first trial after a switch, even when there is ample time to prepare for a switch (e.g., Allport & Wylie, 2000; De Jong, 2000; Gopher, 1996; Goschke, 2000; Meiran, 1996; Rogers & Monsell, 1995). The most compelling evidence for this is the observation that, when switching between two tasks of unequal “strength” (e.g., word reading and color naming), there is sometimes a larger RT cost when switching to the stronger task (e.g., Allport et al., 1994; Allport & Wylie, 2000; Meuter & Allport, 1999). A process of reconfiguring cognitive processes should not take longer for a more familiar task than for a less familiar task. The surprising asymmetry is explicable in terms of competition between, and priming of, task sets (Meuter & Allport, 1999). To succeed in performing the weaker task of a pair (e.g., color naming), one must apply an active control bias toward the weaker task set, inhibiting the stronger task set (word naming) and/or applying extra activation to the weaker task set. When the stronger task set is then required, this negative and/or positive priming must be overcome. In contrast, little control input is needed to ensure performance of the stronger task on a nonswitch trial, so there is little priming to carry over to the next trial when the task changes to the weaker one. This surprising asymmetry of costs turns out to occur only under some conditions (Monsell, Yeung, & Azuma, 2000), suggesting that TSI is only one component of switch cost. That it contributes at all, however, is evidence for priming of a task set over and above any priming of individual S–R associations, because substantial switch costs are observed even when neither the stimulus nor the response are allowed to repeat on adjacent trials (e.g., Allport & Wylie, 2000) and, indeed, when stimuli rarely repeat (Jersild, 1927; Rogers, 1993; Spector & Biederman, 1976).

The second principle we will need is that whole task sets, not just individual responses, may be activated by stimuli. Frontal lobe pathologies, such as “utilization behavior,” in which habitual actions are evoked inappropriately by the mere presence of objects that afford them (Lhermitte, 1983; Shallice, Burgess, Schon, & Baxter,

1989), and everyday action errors, in which our behavior is “captured” by a habitual action associated with a familiar context (Norman, 1981), were a major part of the motivation for Norman and Shallice’s (1986) well-known theory of executive control. The theory supposes that stimuli activate not just responses but learned procedural schemata, which, if suitable biasing control is not applied by a superordinate “supervisory attention system,” may win a “contention scheduling” competition for control of behavior on the basis of their frequency or recency of use alone. However, there is a certain ambiguity in these naturalistic examples about whether it is a task set or a single action that is evoked by the stimulus. Task-switching experiments offer more direct evidence for exogenous activation of task set. Rogers and Monsell (1995) compared congruent stimuli (where both attributes were associated with the same response), incongruent stimuli (the two attributes were associated with opposite responses), and neutral stimuli (the irrelevant attribute was associated with neither response). Neither relevant nor irrelevant attributes were allowed to repeat immediately. It was harder to respond to incongruent stimuli than to congruent stimuli, and this effect was magnified on the first trial after a task switch (hard to explain without some notion of the irrelevant task-set being stronger, or less suppressed, on a switch trial). Also, responses to neutral stimuli were substantially faster than responses to congruent stimuli, especially on switch trials. It is hard to understand why an irrelevant stimulus attribute associated with the correct response should cause more interference than one associated with no response, unless the irrelevant attribute is activating not just responses but also the task set associated with it and thus increasing competition at the task-set level. Similarly, Allport and Wylie (2000) now conceptualize TSI as due to retrieval from memory of a task set (or the inhibition of a task set) recently associated with the stimulus or a stimulus of the same class.

Returning now to the Klein effect, our suggestion is that the presence of a letter string in the stimulus may interfere with the task of color naming not because it activates a single competing response but because it activates the competing task set of word reading—the whole set of associations between orthographic patterns and their meanings or names. The two processing pathways depicted in Figure 2 should be construed as two states of the system in competition, each having an activation level or “readiness” that is a function of control input, practice, priming from recent trials, and the presence of stimulus attributes normally associated with that task. If these attributes are relatively low level perceptual features common to all words, then a pronounceable nonword could activate the task set of reading no less than a word, and the frequency of the word would also be immaterial. This speculation—that the Klein effect would not be modulated by lexical status or frequency—was tested in each of the experiments reported here. The second and third experiments also included item types intermediate in their perceptual properties between nonwords and the control stimuli.

Lexical/Semantic Effects on Color Naming

To challenge the notion that color naming is sensitive to lexical properties of an orthographic stimulus may seem quixotic, given evidence on the effects of three variables: frequency, semantic priming, and emotional salience.

Frequency. Klein (1964) and Fox et al. (1971) reported significantly longer color-naming latencies for “common” words than for “rare” words (Figure 1). However, although Fox et al.’s methodology improved somewhat on Klein’s, it remained less than ideal. There were only five (Klein) or four (Fox et al.) words per word type, each necessarily repeated many times. The “automaticity” of lexical access is not best tested with words whose recognition and pronunciation are strongly primed by frequent repetition. It is also odd to test for a frequency effect with repeated stimuli, given that effects of frequency and repetition typically interact underadditively (see Monsell, 1991, for review). The list-reading method requires that different types of items are presented on different cards, increasing the likelihood of differences in response criterion or arousal between conditions. Finally, list reading allows preview: An overlap between early stages of processing of each item with later stages of processing of the previous item may conceal interference effects.

In our laboratory, Perry and Elliot (1990) manipulated experienced frequency within item by having law and chemistry students color-name technical terms from law (e.g., *tort*) and chemistry (e.g., *isomer*), as well as everyday words matched for their (low) frequency of usage in ordinary discourse, all presented once only per subject. Color-naming latencies were not longer for the category of terms with which the students were familiar. A similar null result was reported by Mogg and Marden (1990), although an effect of expertise on interference has been reported by Dalgleish (1995). Hence, a more direct test is needed of the common assumption, apparently supported by the data of Klein (1964) and Fox et al. (1971), that color naming of word forms is sensitive to their frequency.

Semantic priming. Psycholinguists have used the Klein effect to measure semantic activation, especially for ambiguous words. For example, a spoken word, usually in a sentence context, may be followed by a visual probe word whose color the subject must name. Slower naming of the color of the visual word when it is semantically related to the preceding auditory word is interpreted as an index of the meaning(s) activated by the auditory word (e.g., Conrad, 1974; Jones, 1989; Merrill, Sperber, & McCauley, 1981; Oden & Spira, 1983; Whitney, McKay, Kellas, & Emerson, 1985). This may seem to imply that there is unavoidable lexical access to the probe’s meaning when color naming. However, these semantic priming effects do not rule out the possibility that, during color naming, successful lexical access occurs only for words primed by congruence with the linguistic context or only when the person has just switched from attending to the linguistic content of earlier words.

Emotional salience. The Klein effect has also been used as an index of psychopathology, under the label of

the *emotional Stroop* effect. As just one example, anorexic and bulimic patients name the colors of words relating to food, eating, body shape, and body weight more slowly, relative to emotionally neutral words, than do control subjects (e.g., Jones Chesters, Monsell, & Cooper, 1998, who cite similar studies on a wide range of pathologies). In a theoretical review of the emotional Stroop effect, Williams, Mathews, and MacLeod (1996) elaborate the common view that the effect represents a form of “attentional bias” toward items related to concepts of current concern. They use Cohen et al.’s (1990) connectionist framework, suggesting that (1) relevance of concepts to current concerns raises the activation level of associated lexical input units, (2) a history of association with threat or loss augments the activation of lexical units through a form of neuromodulatory control, and (3) prolonged attention to and rumination on particular ideas increase the strength of stimulus-to-meaning associations. They cite Klein (1964) as a demonstration of the effect of frequency to support the last conjecture, given the conflicting evidence on the effect of expertise already mentioned.

The standard interpretation of the emotional Stroop effect thus takes for granted involuntary lexical access to a word that a person is attempting to color-name. Again, however, it is possible that when one is engaged in color naming and trying to ignore the content, involuntary lexical access occurs (enough to cause interference) only for words primed by the high emotional salience of associated concepts or only under circumstances when the subject has been alerted by other stimuli to the probable occurrence of disorder-salient words.

In conclusion, although the apparent modulation of the Klein effect by frequency, semantic priming, and emotional salience has been taken to implicate involuntary lexical access in interference with color naming, existing evidence is compatible with an alternative possibility: When the task set of color naming is engaged, only words primed by repetition, by semantic congruence, by high temporary or permanent salience, or by membership of the response set tend to evoke a specific associated name (and/or meaning) that competes for production with the color name.

EXPERIMENT 1

In Experiment 1, we compared the interference with color naming produced by words in three frequency bands and by pseudowords, matching for other properties. The effect of frequency on word naming is smaller for regular words (Monsell et al., 1989; Seidenberg, Waters, Barnes, & Tanenhaus, 1984) because the pronunciation of such items conforms to sublexical spelling-sound regularities and is less dependent on recognition of the whole letter string. We therefore used only words with exceptional spelling-sound correspondences. As control strings, we used strings of nonalphabetic characters (e.g., “@£>##&”) approximately matching the words in pixel density, visual complexity, and heterogeneity. Un-

like Klein (1964) and Fox et al. (1971), we measured discrete RTs and mixed different kinds of items randomly within blocks. To check that the lexical status of the pronounceable items and the frequency of the words had the expected effect on the generation of pronunciations, we also measured naming latency for the same items. To avoid any repetition priming, each item was seen just once by each subject, either in the item-naming part or in the color-naming part of the experiment.

Method

Subjects. Twenty undergraduates at the University of Cambridge participated as subjects without pay, but they were encouraged by a small prize for the best combination of short average latency and few errors.

Materials. There were 80 high-frequency (>50 per million) words, 80 medium-frequency (14–35 per million) words, 80 low-frequency (1–4 per million) words, and 80 pronounceable pseudo-words (see Appendix A). (The frequencies are from Kučera and Francis, 1967.) All items were monosyllabic or disyllabic and began with a voiced onset to minimize onset detection problems. Sets were matched in mean number of characters, mean number of syllables, and initial sounds. Word sets were also roughly matched in the proportion of verbs, adjectives, concrete nouns, and abstract nouns, and in the stress pattern of disyllabic items. The words were exception words: A reader who did not know one of these items would be likely to pronounce it wrong. Each set was subdivided into two lists of 40, approximately matched on these properties. Forty control strings of nonalphabetic characters were constructed from the symbol set @, £, \$, #, &, >, <, ?, *, %, and ||, mimicking the words in the heterogeneity and occasional repetition of symbols.

Design. The subjects performed the color-naming and item-naming tasks in a balanced order, naming items from one list, and color-naming items from the other list together with the control strings. There were 5 subjects for each combination of task order and assignment of list to task.

For color naming, 8 items from each set were randomly assigned to trial blocks, and their order was randomized anew, for each subject. Over each group of 5 subjects, every item occurred once in each of the colors: red, green, yellow, blue, and purple. For item naming, 10 items from each set were randomly assigned to blocks, and their order was randomized. For both tasks, 3 warm-up trials using practice items began each block.

Procedure. Before the color-naming task, the subjects were shown samples of the five colors and were told what name should be used. For color naming, items were displayed on a dark screen in one of the five colors. For item naming, items were displayed in white. The screen of the CUB color monitor was located 50 cm from the subject's eyes and controlled by a BBC Model B computer. A throat microphone connected to the computer via a voice key detected responses. RTs were recorded to the nearest 10 msec (the timing resolution of the computer's response interface). Each task was preceded by a block of 15 practice trials

On each trial, a string of white plus signs, equal in length to the upcoming stimulus, was displayed in the center of the screen for 1 sec and then replaced by the stimulus until a response was detected. An interval of 2 sec then elapsed before the next fixation display. The subjects were asked to respond as quickly as possible, avoiding errors, and were told their mean RTs and error rates at the end of each block. An experimenter monitored the voice key's operation and coded errors.

Results

We excluded from analysis all trials on two nonwords that were especially difficult (*oice* and *hirlch*) and four

color or color-associated words inadvertently included in the word sets (*beige*, *blood*, *khaki*, and *color*). Mean correct RTs (excluding a few dubious detections of speech onset) and error rates for each condition are summarized in Figure 3. (Effects on median correct RTs were also analyzed but differed in no interesting way from effects on means.)

Item naming. High-frequency words were named 55 msec faster than low-frequency words, with an intermediate latency for medium-frequency words. Nonwords were named 29 msec slower than low-frequency words. This effect of item type on naming latency was highly reliable [$F_1(3,48) = 77.9, p < .001$; $F_2(3,306) = 49.7, p < .001$], with all pairwise differences significant ($p < .05$) in Newman–Keuls analyses. A similar ordering was visible for the error rates for words [$F_1(2,32) = 18.3, p < .001$; $F_2(2,230) = 12.1, p < .001$], with a reliable pairwise contrast between medium- and low-frequency words ($p < .01$).¹ (We cannot meaningfully compare error rates for words and nonwords: For each word, there was a single correct pronunciation, but, for each nonword, there was some arbitrariness in what was counted as correct.)

Color naming. There was a reliable effect of item type on latency [$F_1(4,64) = 19.4, p < .001$; $F_2(4,384) = 15.2, p < .001$], but there was no sign of the gradient of interference predicted by an associative strength theory. Newman–Keuls contrasts showed that the control strings were color-named faster than any other kind ($p < .01$, by subjects and by items). The only other pairwise difference to reach significance was that, in the by-subjects analysis, high-frequency words were color-named slightly faster than all other types of pronounceable item ($p < .05$). The effect of item type did not quite reach significance in a by-items analysis of variance (ANOVA) excluding the symbol strings [$F_2(3,306) = 2.28, p = .08$]. There was a reliable effect on latency of the color named [$F_1(4,64) = 35.6, p < .001$], with the ordering red < yellow = blue < green < purple. The effects of color and word type interacted [$F_1(16, 256) = 2.55, p < .01$], with the effect of word type somewhat smaller for the colors named more rapidly. Equivalent analyses of error rates yielded no reliable effects of item type ($F_s < 1$).

As a further indicator of the independence of color-naming and item-naming latencies, there was no reliable correlation between the mean color-naming RT (averaged over subjects) for the pronounceable items and their item-naming RT [$r(313) = .025$].

Discussion

Frequency and lexical status had strong effects on item-naming latency, in the direction one would expect from differences in strength of experienced association between visual form and name. The pronounceable items of every type interfered with color naming (relative to the nonalphabetic strings), but there was no sign of a gradient of interference predicted by strength of association. Pseudowords generated just as much interference as words, and latency for high-frequency words was, if

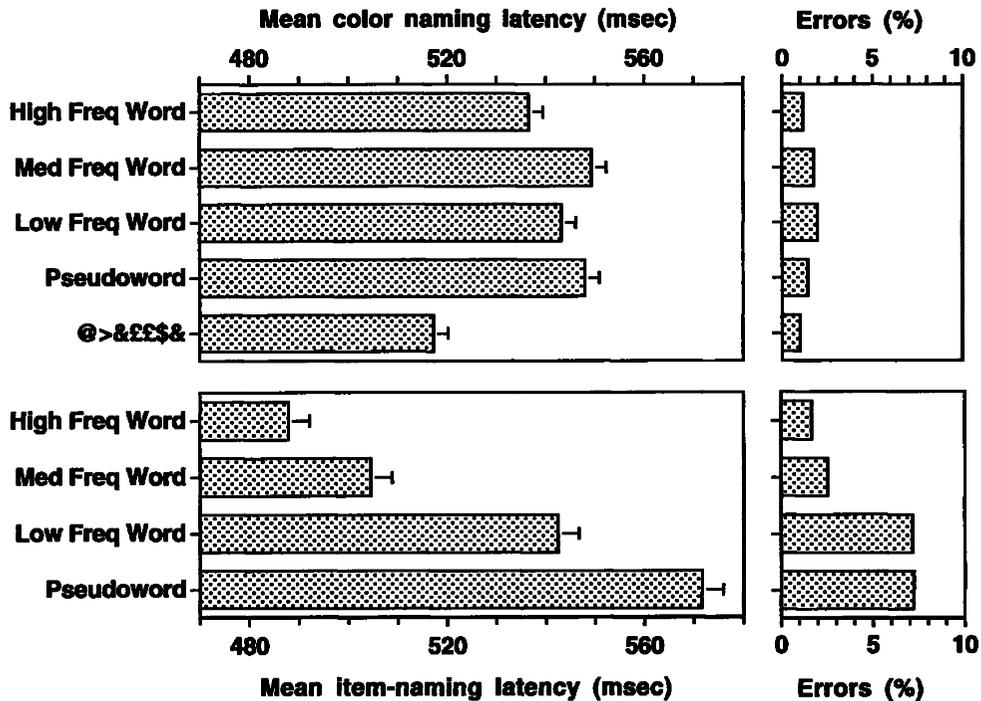


Figure 3. Mean correct response time and standard error (in milliseconds) and error rate (%) for color naming (upper panels) and item-naming (lower panels) in Experiment 1.

anything, slightly shorter than that for medium- or low-frequency words. This pattern of results would appear to be incompatible with the claim that the interference arises from involuntary activation of the pronunciation (or meaning) of the item, with the amount of interference depending on strength of association between the individual letter string and its name or meaning.

EXPERIMENT 2

In Experiment 2, we compared color-naming performance for high-frequency words, low-frequency words, pronounceable pseudowords, two kinds of nonpronounceable letter string, and control strings. In this experiment, we made even greater efforts to match item sets on their nonlexical properties than we did in Experiment 1. The pronounceable items were matched not only with respect to length, onset, stress pattern, and syntactic class, as in Experiment 1, but also with respect to Coltheart's *N*—a measure of the number of lexical neighbors (Coltheart, Davelaar, Jonasson, & Besner, 1977). The pseudowords were formed by recombining the onsets and bodies of the words, to minimize unintended differences between nonwords and words. As control strings, we now used “false font” strings—strings constructed from 26 novel letter-like hieroglyphs matched in number of pixels to, and systematically substituted for, letters in words.

The other main difference from Experiment 1 was that two other types of item were included to test the notion that the tendency for a stimulus to evoke the task set of

reading, and, hence, the degree of interference, might depend on how word-like its sublexical attributes are. We used unpronounceable heterogeneous letter strings (either consonants only or vowels only) and homogeneous strings consisting of one consonant or vowel repeated (like the control strings of Fox et al., 1971).

Method

Materials. There were sets of 80 high-frequency words, 80 low-frequency words, 80 pronounceable pseudowords (each set divided into matched lists of 40—List 1 and List 2, as in Appendix B), 40 heterogeneous letter strings, 40 homogenous letter strings, and 40 false font strings.

The high-frequency words had CELEX frequencies (Burnage, 1988) of at least 30 per million in modern British English, with a geometric mean of 119 per million. The low-frequency words had frequencies in the range 2–5 per million. Pseudowords were constructed by recombining onsets and bodies of the words. As shown in Table 1, the three sets of pronounceable items were matched for Coltheart's *N* (the number of other words that can be formed by changing just one letter), number of letters, number of phonemes, number of syllables (66% monosyllabic, 34% disyllabic), and onset category (open vowel, unvoiced fricative, etc.). (Mean *N* was well matched, but it was impossible to achieve quite the same spread of *N* values for pseudowords as for real words.) High- and low-frequency words were additionally matched for distribution over syntactic class and the syllabic stress of the disyllabic items.

For each letter of the display font, a false font character was created with the same number of pixels and a similar set of straight lines, curves, and closed loops, but not resembling a familiar letter. Forty control items were created by substituting false font characters for their equivalent letters in the List 1 pseudowords. The 40 heterogeneous strings were randomly generated, 15 using vowels only (e.g., OEEI) and 25 using consonants only (e.g., LWLB), roughly

Table 1
Properties of the Word Sets Used in Experiments 2 and 3 (40 Words Per Set)

	Coltheart <i>N</i>	Number of Letters	Number of Phonemes	Ln (Frequency)	Syntactic Class				Onset Class						
					Noun	Verb	Adjective	Other	OV	UF	AP	VS	US	N	VF
High-Frequency Words															
List 1	7.4 (8.3)	5.1 (1.1)	3.9 (1.1)	4.7 (1.0)	27	5	6	2	4	8	12	7	7	2	0
List 2	7.7 (8.6)	5.1 (1.0)	4.0 (1.0)	4.9 (1.0)	31	3	5	1	2	7	12	7	9	3	0
Low-Frequency Words															
List 1	7.5 (7.5)	4.9 (1.1)	4.0 (1.1)	0.7 (0.6)	26	6	7	1	2	11	12	5	7	2	1
List 2	7.7 (10.0)	4.9 (1.0)	3.8 (1.0)	0.7 (0.7)	28	5	5	1	2	6	12	8	9	2	1
Pseudowords															
List 1	7.2 (5.4)	4.7 (1.0)	3.9 (0.9)												
List 2	7.6 (5.6)	4.7 (1.0)	3.8 (1.2)												

Note—The values shown are mean (standard deviation) for the first four columns, counts for the remainder. Frequency is CELEX lexeme frequency, per million. OV, open vowel; UF, unvoiced fricative; AP, approximant (e.g., /r/, /w/); VS, voiced stop; US, unvoiced stop; N, nasal; VF, voiced fricative.

matching the frequencies of letter occurrence and number of repeated letters in words and their distribution of lengths. The homogenous strings similarly matched the words in length distribution and (roughly) letter frequency, with 15 constructed from a vowel (e.g., UUUU) and 25 from a consonant (e.g., WWWW).

Subjects, Design, and Procedure. After replacing 4 slow or error-prone subjects, there were 24 new paid subjects from the same population as for Experiment 1. Each did both tasks but saw each item only once; counterbalancing of task order and assignment of lists to task was the same as in Experiment 1, with 6 subjects per combination. Item-naming trials were divided into 3 blocks, and color-naming trials were divided into 6 blocks, each of 40 experimental trials preceded by 8 warm-up trials. Four colors were used: red, green, blue, and purple, randomly paired with items, subject to the constraint that, over all subjects, each color was paired equally often with each item, and, within subjects, each color was used equally often with each item type. The order of items was randomized anew for each subject in such a way that each block contained an approximately equal number of items of each type. The item-naming task was preceded by a 9-trial practice block, and the color-naming task was preceded by an 18-trial practice block. The procedure was otherwise identical to that of Experiment 1.

Results

Mean correct RTs (excluding dubious onset measures) and error rates are shown in Figure 4.

Item naming. Frequency and lexical status again had the effects expected on item-naming latency [$F_1(2,40) = 87.7, p < .001$; $F_2(2,234) = 51.3, p < .001$], with all pairwise contrasts reliable ($p < .01$). High-frequency words were also named more accurately than low-frequency words [$F_1(1,20) = 27.3, p < .001$; $F_2(1,156) = 27.7, p < .001$]. The error rate for nonwords was intermediate but, as noted above, was not exactly comparable to that for words.

Color naming. There was a significant effect of item type on mean correct RT [$F_1(5,100) = 15.17, p < .001$; $F_2(5,468) = 14.99, p < .001$]. Pairwise comparisons ($p < .05$) show that (1) false font strings were color-named significantly faster than all other types both by subjects and by items, (2) homogenous and heterogeneous letter strings were named significantly faster than pseudowords or low-frequency words both by subjects and by items, and (3) high-frequency words were color-named significantly faster than low-frequency words by

subjects and faster than both low-frequency words and pseudowords by items. The effect on error rate was not reliable [$F_1(5,100) = 1.59, p = .17$; $F_2(5,468) = 1.95, p = .08$], although the false font strings were named most accurately. Although high-frequency words were color-named faster than low-frequency words, there were a few more errors on the high-frequency words.

Color-naming RT was influenced by the color (red < green = blue < purple) [$F_1(3,60) = 34.6, p < .001$], but this did not interact reliably with item type. Error rate was not reliably influenced by color. Heterogeneous letter strings consisting of vowels might be thought more pronounceable than those consisting of consonants, but the consonant strings were color-named only 1.7 msec slower than the vowel strings [$t_2(38) = 0.34$].

As a further test of the independence of color-naming and item naming latencies, we regressed the average color-naming time for each item against its word-naming latency; the correlation was trivial ($r = -.014$).

Discussion

The results of Experiment 2 confirm those of Experiment 1. Although substantial and reliable interference with color naming was observed for high- and low-frequency words and pseudowords (34 msec on average), there was no gradient of interference corresponding to the ease of naming these items. The average pseudoword was color-named no faster than the average word, and the average high-frequency word was actually color-named significantly faster than the average low-frequency word (but a little less accurately). The amount of interference with color naming did not correlate negatively with item-naming latency.

Again, this outcome disconfirms any expectation that the amount of interference with color naming will depend on the strength of association between a stimulus and its particular pronunciation, as indexed by the ease of naming the item. However, we did see a gradient of interference between the pronounceable pseudowords and the false font strings, with heterogeneous but unpronounceable strings color-named faster than pseudowords and significantly faster than low-frequency words, and homo-

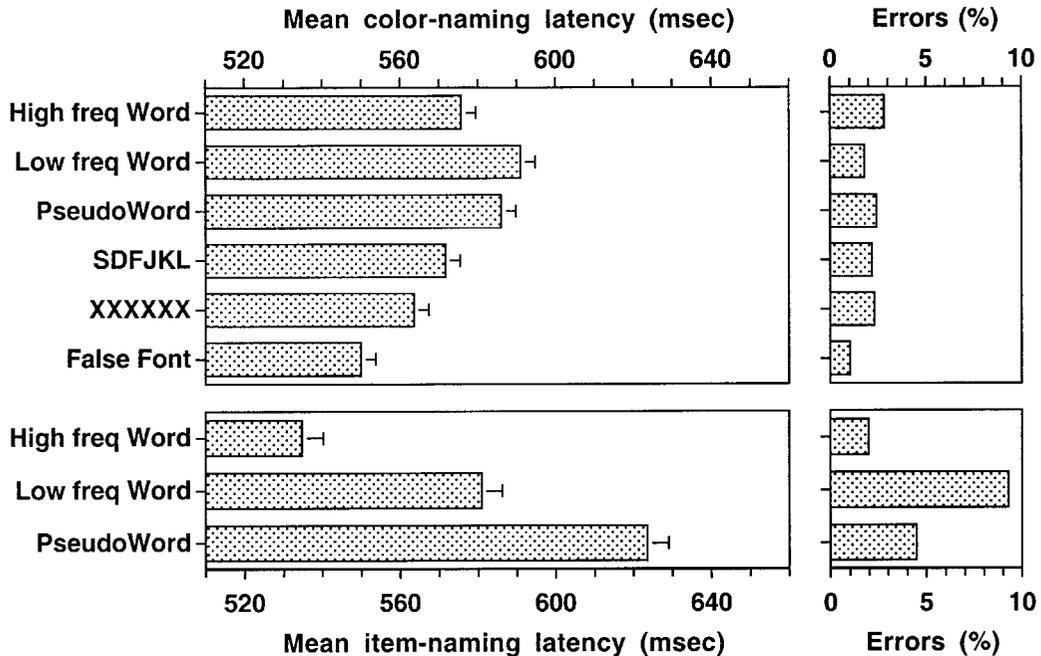


Figure 4. Mean correct response time and standard error (in milliseconds) and error rate (%) for color naming (upper panels) and item naming (lower panels) in Experiment 2.

geneous strings faster still, but still significantly slower than false font strings. This observation suggests that the degree of interference is determined by how word-like low-level perceptual attributes of the stimulus are, not by contact with a specific lexical representation. This outcome favors the hypothesis that the interference results not from competition from a tendency to say the word's name (a tendency whose strength should be indexed by naming latency) but from competition at the task-set level, due to attributes of the stimulus activating the inappropriate task set associated with them.

EXPERIMENT 3

Experiments 1 and 2 mixed stimulus types within a block. The experiments of Klein (1964) and Fox et al. (1971) did not mix stimulus types within a list. Blocking the types permits subjects to vary their response criterion as a function of stimulus type and, hence, seems methodologically undesirable. However, we conducted Experiment 3 to see whether the lack of an effect of lexicality on color naming and the unexpected direction of the frequency effect in Experiments 1 and 2 would survive blocked presentation.

Method

We tested a new set of 24 subjects from the same population (replacing 2 who exceeded the average time and error criteria). The only difference from Experiment 2 was that the item type was constant within a block. The order of item types was determined, for each combination of task order and list assignment, by a balanced Latin square.

Results

The data are plotted in Figure 5. Comparison with Figure 4 suggests that, though blocking had quite a substantial effect on the ordering of mean RTs in the item-naming conditions, it had relatively little effect on the ordering of mean RTs in the color-naming conditions.

Item naming. The effect of item type on mean correct RT was reliable [$F_1(2,40) = 69.4, p < .001$; $F_2(2,234) = 97.7, p < .001$], and pairwise contrasts showed that high-frequency words were named faster ($p < .01$) than low-frequency words and pseudowords, which did not differ. The error rate difference between high- and low-frequency words was also reliable [$F_1(1,20) = 45.9, p < .001$; $F_2(1,156) = 34.5, p < .001$]. Analyses combining data from Experiments 2 and 3 show that the interaction between experiment and item type was reliable for RT [$F_1(2,80) = 17.9, p < .001$; $F_2(2,234) = 52.1, p < .001$], but not for errors. The error rate on words was, however, significantly lower for the blocked experiment [$F_1(1,40) = 4.15, p = .048$; $F_2(1,156) = 59.7, p = .002$].

Color naming. With blocked presentation, there were significant main effects on mean correct color-naming RT of item type [$F_1(5,100) = 8.0, p < .001$; $F_2(5,468) = 30.4, p < .001$], which did not interact reliably with the effect of color ($F_1 = 1.3$). Newman-Keuls pairwise comparisons, by subjects and by items, show that the three pronounceable types of item were color-named significantly slower ($p < .01$) than false font strings (by 40 msec on average) but the differences among them were not reliable. Heterogeneous letter strings were color-named faster than any of the pronounceable types ($p < .05$ by subjects; $p <$

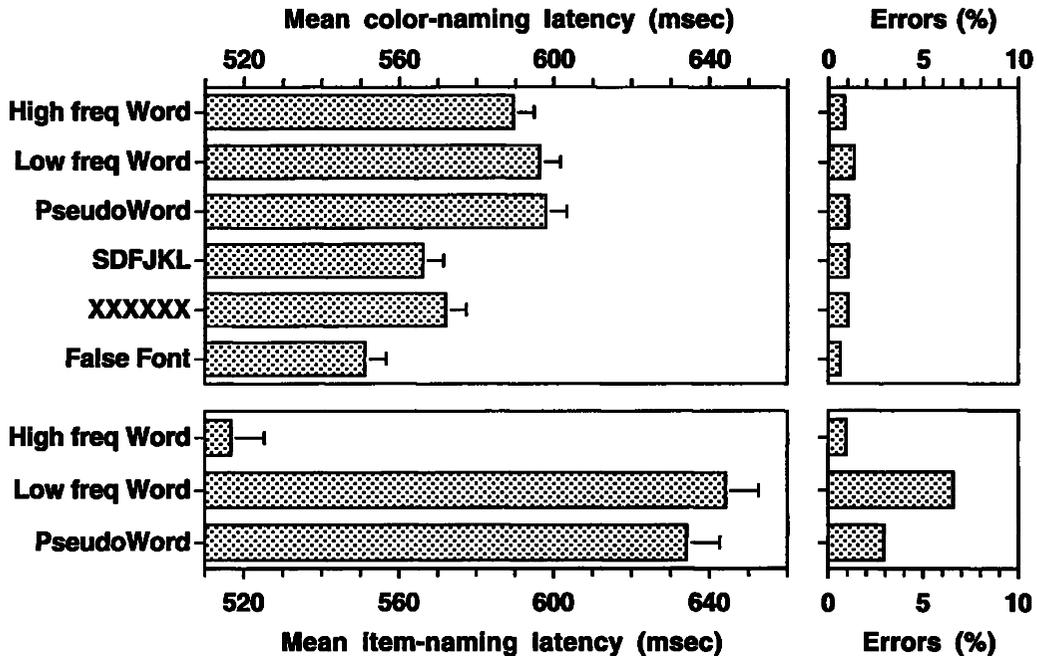


Figure 5. Mean correct response time and standard error (in milliseconds) and error rate (%) for color naming (upper panels) and item naming (lower panels) in Experiment 3.

.01 by items) but reliably slower than false font strings only in the by-items analysis ($p < .01$). Color-naming latencies for homogenous letter strings were reliably longer than those for false font strings only by items ($p < .01$), shorter than those for pseudowords and low-frequency words by subjects ($p < .05$), and shorter than for all three pronounceable types by items ($p < .01$). Error rates did not differ reliably among item types ($F_s < 1$). An analysis of RTs for the heterogeneous strings showed that consonant strings were color-named only 6.7 msec slower than vowel strings [$t_2(38) = 0.94$].

Color-naming times for the pronounceable items were regressed against item-naming times. Although statistically reliable [$r(239) = .16, p < .05$], the correlation was small and positive—the unexpected direction. As an indicator of the potential power of these correlations, given that different subsets of subjects supply the color- and item-naming time for each item, we note the correlation between item-naming times in Experiments 2 and 3: $r(239) = .698, p < .0001$.

In a combined analysis of correct mean color-naming RT from Experiments 2 and 3, the interaction between experiment and item type was reliable only in the by-items analysis [$F_1 < 1$; $F_2(5,468) = 2.32, p = .04$]. There was no reliable interaction between experiment and item type in the error rates ($F_s < 1$), though the overall error rate was reliably lower in Experiment 3 [$F_1(1,40) = 9.35, p = .004$; $F_2(1,468) = 23.36, p < .001$].

As a further test of the effects of frequency, we regressed the by-items color-naming RT within each of the two frequency bands, combined over experiments, against

N , log frequency, number of letters, and number of phonemes. No reliable effects were found.

Discussion

In one sense, Experiment 3 confirms the methodological inadequacy of blocked presentation: In the item-naming part of the experiment, the subjects appeared to adjust their response criterion as a function of the material type, so that the frequency effect appeared larger and mean RT for low-frequency exception words became greater than that for pseudowords. Such effects on naming latency of “context” (i.e., the frequency, lexical status, and regularity of the other items named) are well known and have been attributed to changes in response criterion and other adjustments (Lupker, Brown, & Colombo, 1997; Monsell, Patterson, Graham, Hughes, & Milroy, 1992).

It may be reasonable for subjects to name more cautiously when expecting only low-frequency exception words, but it would be more surprising if they significantly adjusted their color-naming strategy as a function of the item type. Indeed, the effects of item type on color naming were very similar with mixed and blocked presentation. Hence, Experiment 3 indicates the robustness of the main outcome of Experiments 1 and 2. For pronounceable strings, the interference with color naming was uninfluenced by lexical status, and the interference was no greater for high-frequency words than for low-frequency words, even though frequency had a substantial effect on item-naming time. (If anything, high-frequency items were again easier to color-name, and there was a

slight positive, rather than negative, correlation over items between item- and color-naming RTs.) As before, pronounceable items containing letters were color-named faster than pronounceable strings but slower than false font strings.

GENERAL DISCUSSION

In three experiments, subjects named the color in which a pronounceable word or pseudoword was displayed, each item being seen only once. Color naming was between 20 and 40 msec slower (and slightly less accurate) for such an item than for one that did not contain letters but was visually equivalent, in the sense of having the same number of pixels, characters, and the same heterogeneity of forms. We have called this type of interference the *Klein effect*. If it arose as a result of competition between individual responses—a tendency to name the word rather than the color—we would expect the Klein effect to be greater for items that are more efficiently named. Both lexical status and frequency had large effects, in the expected direction, on naming latency. The surprising but robust outcome of our experiments is that the interference with color naming was no greater for words than for pseudowords and was no greater for high-frequency words than for words with much lower frequency. Nor did the Klein effect observed for an item show an inverse correlation over items with naming latency for that item.

If the interference were caused by involuntary lexical access and retrieval/computation of the item's pronunciation (and/or meaning), there should have been more interference from familiar words than for less familiar words and more interference from a known letter string than from a novel but pronounceable letter string. In S–R terms, the interference with color naming did not depend on the strength of association between individual stimulus and individual response, as would be predicted on the basis of a model of the class exemplified by Cohen et al. (1990).

What then is the source of the Klein effect? And what is the relationship between the Klein effect and the much larger interference effect typically observed from color words (the classic Stroop effect) or from words of high emotional or contextual salience? Ours is essentially a two-factor account. We propose that there is competition for control of behavior, first, at the task-set level between whole S–R mappings and, second, at the level of individual responses from any inappropriate response tendency that succeeds in being activated by an irrelevant stimulus or attribute in spite of the current level of task-set activation.

Competition Between Task Sets

In the introduction, we drew attention to task-switching research that suggests that task sets may be construed as having degrees of activation or “readiness,” with the level of activation being a function of control input or bias (the intention to perform one task rather than another), task

strength (greater for practiced tasks), priming from the task performed on previous trials (the “task-set inertia” of Allport et al., 1994), and the strength of association between stimulus properties and the task set. When the subjects color-named, a control bias was endogenously applied on almost all trials sufficient to enable the relevant task set (the mapping from colors to their names) and suppress the stronger (because more practiced) task set of reading. However, we assume that the detection of word-like properties of the stimulus evokes exogenously in literate subjects the associated task set of reading. That is, although executive processes are trying to, “turn off” the word-reading task set, the readability of each stimulus tends to turn it back on. We propose that this partial activation of the reading task set, although usually insufficient for an unprimed word to activate its associated name and meaning, competes with activation of the color-naming task set and impairs the efficiency of translation between color and name.

This competition between task sets might be implemented in the Cohen et al. (1990) architecture as follows. First, instead of having the task nodes “on” or “off,” they could have graded activation with a fixed quantity of activation divided between them. Second, there would have to be a sublexical layer of “perceptual nodes” to detect word-like properties and modifiable connections from the perceptual to the task nodes. Endogenous activation of the reading task node in the presence of sublexical attributes characteristic of words would strengthen connections between the perceptual nodes activated by those attributes and the task node, creating a source of exogenous activation of the reading task set and thus making reading more “automatic.”

What properties of a letter string might determine whether it is “word-like”? We observed in Experiments 2 and 3 an intermediate degree of interference when the stimulus contained letters but was unpronounceable. Containing letters might therefore be one easily detected property diagnostic of word-likeness. In the (methodologically superior) Experiment 2, there was less (though not reliably less) interference for a homogenous letter string than for a heterogeneous letter string, suggesting that the critical perceptual properties may include, in addition, some heterogeneity of letter sequence. The greater interference for pseudowords than for unpronounceable letter strings requires an additional property. It could be pronounceability, detected by the rapid activation of sublexical orthographic to phonological connections (cf. Van Orden & Goldinger, 1994). Or it could be the detection of patterns with consonants and vowels arranged in a orthographic syllabic structure (cf. Spoehr & Smith, 1973; Taft, 1979, 1987). Neuroimaging researchers have identified areas of cortex (both “visual” and “phonological”) that are equally excited by words and pseudowords, and much more so than by consonant or false font strings (e.g., Price, Wise, & Frackowiak, 1996)—potential neural correlates of detection of word-likeness in the sense that we need.

Competition Between Response Tendencies

The construct of task set may, in the present context, be construed as (in part) a filter, a gate, or a threshold designed to prevent a response tendency associated with a currently irrelevant stimulus attribute from being activated. There is a clear parallel between this function of a task set and “filter” or “spotlight” early selection theories of auditory and visual attention. For example, in Treisman’s (1960) filter-attenuation early selection theory, the filtering is a partial attenuation of the information from unattended sources. Treisman proposed that a word permanently primed by permanent salience or transiently primed by contextual appropriateness needs less sensory input for recognition and can thus exceed a threshold for identification even when the input is attenuated. In his recent review of the early/late selection debate, Pashler (1998) concluded that this account remains compatible with more recent data indicative of “breakthrough” of meaning from an unattended source. Along the same lines, we propose that, although endogenous suppression of the strong task set of reading, in favor of color-naming, is sufficient to ensure that the color is (usually) named, it will not be sufficient to prevent frequent “breakthrough” of an irrelevant word to response activation when lexical access for the word is sufficiently primed. If a conflicting response tendency is activated in this way by the irrelevant attribute, there will be additional competition to be resolved by a response-selection process. If the same response tendency is activated by both tasks, there will be facilitation. Such response competition or reinforcement effects may be captured by a simple decision mechanism in which the response is chosen whose activation first exceeds that of others by a criterial difference (Cohen et al., 1990), or responses inhibit each other, and the first one to reach an absolute threshold wins (Cohen & Huston, 1994).

In the classic Stroop experiment, with words that are color names, lexical access is primed both by membership of the response set for the color-naming task and by massive repetition of the color names. In other conditions of Klein (1964) and Fox et al. (1971), lexical access for the irrelevant words was presumably primed by a categorical or associative relationship to a word in the response set (and by repetition). In the psycholinguistic uses of the Klein effect described in the introduction, lexical access is primed by contextual salience on trials in which the word is congruent with the context, and the task set of comprehension is presumably also primed by the need to understand the immediately preceding sentence fragment. In the emotional Stroop studies, lexical access for certain words is permanently primed by the high emotional salience of their concepts and this effect may be augmented by emotional arousal (see Jones Chesters et al., 1998). Interestingly, MacLeod (1996) found only an unreliable 4-msec increase in the Klein effect for neutral words primed only by a single 1.5-sec “study” encounter, over a lag of 6–20 sec occupied by other items. Clearly,

the degree and type of priming necessary for significant breakthrough has yet to be established.

The idea of “breakthrough” to response activation may also account for a puzzling feature of our data: the slightly shorter color-naming times observed in all three experiments for high-frequency words than for low-frequency words. We hesitate to make much of this effect: It was small, its reliability was marginal, and, in Experiment 2, there was an opposed error difference. However, we might speculate that, even with unprimed words, on some small proportion of the trials (independent of lexical status or frequency), the subject failed to suppress the task set of reading sufficiently, and the word’s name was activated. The response selection stage then found itself in possession of a name not in the response set and had to reinitiate the generation of, or otherwise “find,” an activated color name. This occasional generation of an inappropriate response would happen more quickly for high-frequency words and be sooner dealt with.

Why, if we are willing to attribute some instances of interference with color naming to a “late” response-selection process, do we not so attribute all such interference? As soon as we attribute all the interference to competition among activated response tendencies, we have trouble explaining why the response selection process does not, under the conditions of our experiments, have more trouble rejecting names that must have been activated more strongly/quickly, as a function of their lexicality or high frequency. Second, any “late-selection” theory inherits the problem endemic to its class (cf. Deutsch & Deutsch, 1963), of specifying how the late selection process knows which is the right response among those activated. It would seem necessary for each activated response tendency to be tagged with, or linked to, its sensory source of activation, in a way that can be “inspected” by the response selection process. The virtue of an early selection mechanism is that much of the work of action selection is done at the perceptual level. On occasions when multiple response tendencies are nevertheless activated, the uncertainty can usually be resolved merely by selecting the stronger or quicker of the activated response tendencies, without reference to its source.

Finally, we note that there is evidence from the Stroop paradigm that the degree of suppression of the reading task set—the efficiency of gating—is a function of the exact nature of the color-processing task. Besner, Stolz, and Boutilier (1997) found that requiring the subject to classify the color of a single colored letter in the word, rather than the whole word, reduced or eliminated Stroop interference. This manipulation allows subjects to narrow the spatial focus of processing (LaBerge, 1983) to help suppress the reading task set; the Stroop effect is also well known to be diluted by spatial separation of color and word (Kahneman & Chajczyk, 1983). The Stroop effect is also reduced by asking the subject to detect, rather than classify, the color (Bauer & Besner, 1997). There are

doubtless numerous other determinants of the efficiency of this gating that we have not addressed here, such as variability in the irrelevant attribute (Melara & Mounts, 1993).

Summary

We propose that when a stimulus affords multiple responses, as with a colored word, there may be two sources of interference with the performance of the weaker task, color naming. The first is competition at the level of whole task sets. If the stimulus contains word-like perceptual properties, this activates the whole task set of reading, and competition from this irrelevant task set reduces the efficiency with which color naming is performed. The second is competition from a specific response tendency, the word's name, activated in spite of the intended suppression of the reading task set. With unprimed words, we propose that only the first source of interference is significant; suppression of the reading task set is generally sufficient to prevent lexical access and consequent interference at the level of individual response tendencies, so that the existence of a lexical entry and the efficiency with which it can be accessed during naming or comprehension do not determine the amount of interference in the Klein effect. However, when lexical access is primed, as in the classic Stroop effect, or, for emotionally salient or contextually congruent words, there is breakthrough to lexical access, both sources of interference contribute, and a larger interference effect is typically observed.

REFERENCES

- ALLPORT, D. A., STYLES, E. A., & HSIEH, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 421-452). Cambridge, MA: MIT Press.
- ALLPORT, D. A., & WYLIE, G. (2000). Task-switching, stimulus-response bindings and negative priming. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 35-70). Cambridge, MA: MIT Press.
- BALOTA, D. A., & CHUMBLEY, J. I. (1985). The locus of frequency effects in the pronunciation task: Lexical access and/or production. *Journal of Memory & Language*, **24**, 89-106.
- BAUER, B., & BESNER, D. (1997). Processing in the Stroop task: Mental set as a determinant of performance. *Canadian Journal of Psychology*, **51**, 61-68.
- BESNER, B., STOLZ, J. A., & BOUTILIER, C. (1997). The Stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review*, **4**, 221-225.
- BOROWSKY, R., & BESNER, D. (1993). Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 815-840.
- BURNAGE, G. (1988). *CELEX: A guide for users*. Nijmegen, The Netherlands: Centre for Lexical Information.
- COHEN, J. D., DUNBAR, K., & MCCLELLAND, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, **97**, 332-361.
- COHEN, J. D., & HUSTON, T. A. (1994). Progress in the use of interactive models for understanding attention and performance. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV* (pp. 453-476). Cambridge, MA: MIT Press.
- COHEN, J. D., USHER, M., & MCCLELLAND, J. L. (1998). A PDP approach to set size effects within the Stroop task: Reply to Kanne, Balota, Spieler, and Faust (1998). *Psychological Review*, **105**, 188-194.
- COLTHEART, M., DAVELAAR, E., JONASSON, J. T., & BESNER, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). London: Academic Press.
- CONRAD, C. (1974). Context effects in sentence comprehension: A study of the subjective lexicon. *Memory & Cognition*, **2**, 130-138.
- DALGLEISH, T. (1995). Performance on the emotional Stroop task in groups of anxious, expert and control subjects: A comparison of computer and card presentation formats. *Cognition & Emotion*, **9**, 341-362.
- DE JONG, R. (2000). An intention-activation account of residual switch costs. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 357-376). Cambridge, MA: MIT Press.
- DEUTSCH, J. A., & DEUTSCH, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, **70**, 80-90.
- DULANEY, C. L., & ROGERS, W. A. (1994). Mechanisms underlying reduction in Stroop interference with practice for young and old adults. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 470-484.
- DYER, F. N. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory & Cognition*, **1**, 106-120.
- FOX, L. A., SCHOR, R. E., & STEINMAN, R. J. (1971). Semantic gradients and interference in naming color, spatial direction, and numerosity. *Journal of Experimental Psychology*, **91**, 59-65.
- GLASER, M. D., & GLASER, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception & Performance*, **8**, 875-894.
- GOLDINGER, S. D., AZUMA, T., ABRAMSON, M., & JAIN, P. (1997). Open wide and say "Blah!" Attention dynamics of delayed naming. *Journal of Memory & Language*, **37**, 190-216.
- GOPHER, D. (1996). Attention control: Explorations of the work of an executive controller. *Cognitive Brain Research*, **5**, 23-38.
- GOSCHKE, T. (2000). Intentional reconfiguration and involuntary persistence in task-set switching. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 331-355). Cambridge, MA: MIT Press.
- GRAINGER, J., & JACOBS, A. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, **103**, 518-565.
- JERSILD, A. T. (1927). Mental set and shift. *Archives of Psychology*, **14** (No. 89), 81.
- JONES, J. L. (1989). Multiple access of homonym meanings: An artifact of backward priming? *Journal of Psycholinguistic Research*, **18**, 417-432.
- JONES CHESTERS, M. H., MONSELL, S., & COOPER, P. J. (1998). The disorder-salient Stroop effect as a measure of psychopathology in eating disorders. *International Journal of Eating Disorders*, **24**, 65-82.
- KAHNEMAN, D., & CHAJCZYK, D. (1983). Tests of the automaticity of reading: Dilution of Stroop effects by color-irrelevant stimuli. *Journal of Experimental Psychology: Human Perception & Performance*, **9**, 497-509.
- KANNE, S. M., BALOTA, D. A., SPIELER, D. H., & FAUST, M. E. (1998). Explorations of Cohen, Dunbar, and McClelland's connectionist model of Stroop performance. *Psychological Review*, **105**, 174-187.
- KLEIN, G. S. (1964). Semantic power measured through the interference of words with color-naming. *American Journal of Psychology*, **77**, 576-588.
- KUCERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- LABERGE, D. (1983). The spatial extent of attention to letters and words. *Journal of Experimental Psychology: Human Perception & Performance*, **9**, 371-379.
- LHERMITTE, F. (1983). "Utilization behaviour" and its relation to lesions of the frontal lobes. *Brain*, **106**, 237-255.
- LUPKER, S. J., BROWN, P., & COLOMBO, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **3**, 570-590.
- MACLEOD, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, **109**, 163-203.
- MACLEOD, C. M. (1996). How priming affect speeded implicit tests of

- remembering: Naming colors versus reading words. *Consciousness & Cognition*, **5**, 73-90.
- MACLEOD, C. M. (1998). Training on integrated versus separated Stroop tasks: The progress of interference and facilitation. *Memory & Cognition*, **26**, 201-211.
- MACLEOD, C. M., & DUNBAR, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 126-135.
- MCCLELLAND, J. L., & RUMELHART, D. E. (1981). An interactive-activation model of context effects in letter perception. Part I: An account of basic findings. *Psychological Review*, **88**, 375-405.
- MEIRAN, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 1423-1442.
- MELARA, R. D., & MOUNTS, J. R. W. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. *Memory & Cognition*, **21**, 627-645.
- MERRILL, E. C., SPERBER, R. D., & MCCAULEY, C. (1981). Differences in semantic encoding as a function of reading and comprehension skill. *Memory & Cognition*, **9**, 618-624.
- MEUTER, R. F. L., & ALLPORT, D. A. (1999). Bilingual language-switching in naming: Asymmetrical costs of language selection. *Journal of Memory & Language*, **40**, 25-40.
- MEYER, D. E., & KIERAS, D. E. (1997). A computational theory of executive control processes and human multiple task performance: Part 1. Basic mechanisms. *Psychological Review*, **104**, 3-65.
- MOGG, K., & MARDEN, B. (1990). Processing of emotional information in anxious participants. *British Journal of Clinical Psychology*, **29**, 227-229.
- MONSELL, S. (1990). Frequency effects in lexical tasks: Reply to Balota and Chumbley. *Journal of Experimental Psychology: General*, **119**, 335-339.
- MONSELL, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 148-197). Hillsdale, NJ: Erlbaum.
- MONSELL, S., DOYLE, M. C., & HAGGARD, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, **118**, 43-71.
- MONSELL, S., PATTERSON, K. E., GRAHAM, A., HUGHES, C. H., & MILROY, R. (1992). Lexical and sublexical translation of spelling to sound: Strategic anticipation of lexical status. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 452-467.
- MONSELL, S., YEUNG, N., & AZUMA, R. (2000). Reconfiguration of task set: Is it easier to switch to the weaker task? *Psychological Research*, **63**, 250-264.
- NORMAN, D. A. (1981). Categorization of action slips. *Psychological Review*, **88**, 1-15.
- NORMAN, D. A., & SHALLICE, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self regulation* (Vol. 4, pp. 1-18). New York: Plenum.
- ODEN, G. C., & SPIRA, J. L. (1983). Influence of context on the activation and selection of ambiguous word senses. *Quarterly Journal of Experimental Psychology*, **35A**, 51-64.
- PAAP, K. R., & JOHANSEN, L. S. (1994). The case of the vanishing frequency effect: A retest of the verification model. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 1129-1157.
- PASHLER, H. E. (1998). *The psychology of attention*. Cambridge, MA: MIT Press.
- PERRY, N., & ELLIOT, R. (1990). [Undergraduate research project, University of Cambridge]. Unpublished raw data.
- PLAUT, D. C., MCCLELLAND, J. L., SEIDENBERG, M. S., & PATTERSON, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, **103**, 56-115.
- PLOURDE, C. E., & BESNER, D. (1997). On the locus of the word frequency effect in visual word recognition. *Canadian Journal of Psychology*, **51**, 181-194.
- POSNER, M. I., & SNYDER, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55-85). Hillsdale, NJ: Erlbaum.
- PRICE, C. J., WISE, R. J. S., & FRACKOWIAK, R. S. J. (1996). Demonstrating the implicit processing of visually presented words and pseudowords. *Cerebral Cortex*, **6**, 62-70.
- ROGERS, R. D. (1993). *The costs of switching between cognitive tasks: A performance analysis*. Unpublished doctoral dissertation, University of Cambridge.
- ROGERS, R. D., & MONSELL, S. (1995). The costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, **124**, 207-231.
- SEIDENBERG, M. S., WATERS, G. S., BARNES, M. A., & TANENHAUS, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning & Verbal Behavior*, **23**, 383-404.
- SHALLICE, T., BURGESS, P. W., SCHON, F., & BAXTER, D. M. (1989). The origins of utilization behaviour. *Brain*, **112**, 1587-1598.
- SPECTOR, A., & BIEDERMAN, I. (1976). Mental set and mental shift revisited. *American Journal of Psychology*, **89**, 669-679.
- SPOEHR, K. T., & SMITH, E. E. (1973). The role of syllables in perceptual processing. *Cognitive Psychology*, **5**, 71-89.
- STROOP, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, **18**, 643-662.
- TAFT, M. (1979). Lexical access via an orthographic code: The basic orthographic syllable structure. *Journal of Verbal Learning & Verbal Behavior*, **18**, 21-39.
- TAFT, M. (1987). Morphographic processing: The BOSS re-emerges. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 265-279). Hillsdale, NJ: Erlbaum.
- TREISMAN, A. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, **12**, 242-248.
- VAN ORDEN, G. C., & GOLDINGER, S. D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 1269-1291.
- WHITNEY, P. W., MCKAY, T., KELLAS, G., & EMERSON, W. A. J. (1985). Semantic activation of noun concepts in context. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 126-135.
- WILLIAMS, J. M., MATHEWS, A., & MACLEOD, C. (1996). The emotional Stroop task and psychopathology. *Psychological Bulletin*, **120**, 3-24.

NOTE

1. For the by-items analyses in all three experiments, RTs were pooled over colors. The factor of list was included to exclude from the error term variance due to differences in the performance of the groups of subjects to whom List 1 and List 2 were assigned for a given task. However, there was in fact only one list of nonpronounceable items, which may artificially reduce the error variance in this analysis relative to the nominal degrees of freedom. To guard against the danger of a Type I error (i.e., pairwise comparisons among the item types), the by-items pairwise comparisons among pronounceable items that we report come from analyses restricted to the pronounceable items (with list as a factor) and pairwise comparisons among nonpronounceable item types (in Experiments 2 and 3) to analyses restricted to the nonpronounceable types (with no list factor).

APPENDIX A

Pronounceable Items Used in Experiment 1

High-Frequency Words

List 1. account, address, advice, are, bear, blood, both, dis-ease, doubt, gross, country, couple, create, eye, expect, head, heavy, honour, include, learn, money, mother, nothing, post, prove, process, sample, success, supreme, suggest, figure, friend, father, trouble, various, watch, work, want, result, remain

List 2. amount, attempt, answer, ancient, beauty, build, death, door, direct, colour, concern, come, cover, exist, extreme, hotel, here, idea, league, mind, month, many, people, pretty, police, circle, source, social, spread, support, foot, front, truth, very, water, woman, worry, region, remain, remove

Medium-Frequency Words

List 1. anchor, attract, alarm, aunt, bowl, bold, bull, diet, di-vine, guilt, clerk, caution, collect, exact, exceed, honey, hatred, ocean, intend, laugh, mileage, museum, nude, prayer, pursuit, promote, sweat, secure, select, serum, folk, finance, fruit, tourist, vague, wealth, whiskey, worst, routine, roll

List 2. adjust, alert, argue, angel, bush, behalf, decay, delight, defend, courage, chorus, chaos, cruel, extend, engage, height, healthy, onion, lover, monk, motive, mild, patrol, parade, pre-cise, cease, sewage, subtle, sue, scholar, flood, fault, tongue, vary, warrant, wound, ward, realm, rhythm, resist

Low-Frequency Words

List 1. abode, ague, attest, arctic, blouse, beige, binder, demise, dungeon, granite, canal, chasm, khaki, err, excite, harem, hearty, ordeal, infer, lament, mishap, monarch, nausea, palsy, propel, perplex, sluice, swan, subdue, psalm, famine, florist, fete, tigress, vow, womb, warp, wharf, renown, rouse

List 2. abyss, asthma, awry, ache, bough, beguile, dearth, draught, defunct, cadet, caste, coupon, coerce, extinct, exalt, hearse, homage, idler, lacquer, malice, malign, monsoon, pall, pigeon, profuse, sieve, surfeit, swarm, supine, sedate, feud, forearm, trough, vase, wand, worm, wan, receipt, repute, rogue

Pseudowords

List 1. allum, arint, aggish, appent, bennel, bistle, braif, delth, disson, grelker, creash, cleedle, crompt, eattock, ernell, hensate, hulf, olpith, isser, lurney, menty, mobe, narchie, parmoil, plonnet, praik, slamp, sneaper, switton, slith, fippick, foodle, fream, trinter, voller, wossle, woze, wilnet, rhistle, riddon

List 2. amsutt, arple, amtry, ane, berkin, blean, daper, desto, danden, cronth, cople, clabby, crown, enue, entle, hoid, hirlich, oice, loak, mardler, misp, munker, pabe, pettic, portish, soit, slube, savver, sallen, spolue, fondule, fraith, tiry, voad, wef, wendo, wurder, relk, rhortle, rimbale

APPENDIX B

Pronounceable Items Used in Experiment 2

High-Frequency Words

List 1. aunt, snow, suit, wear, rough, spread, breath, threat, pass, watch, worse, source, eye, worth, move, heart, son, war, give, young, could, both, great, have, friend, build, beauty, danger, heavy, crisis, cover, country, people, suppose, woman, address, police, answer, machine, worry

List 2. come, done, tongue, search, soul, broad, post, heard, touch, break, youth, wood, foot, blood, month, glass, truth, front, love, death, want, child, work, once, world, good, said, hotel, water, island, engine, foreign, colour, couple, money, nature, trouble, weather, wonder, control

Low-Frequency Words

List 1. ache, bough, graft, guile, leapt, monk, puss, quart, rouse, sew, slant, soot, squat, stead, swamp, swan, ton, vase, warp, wolf, worm, yolk, squash, tread, suave, wry, wasp, morose, swarthy, borough, cognac, hybrid, lacquer, awry, lesion, spongy, halter, covet, coupon, bully

List 2. thwart, crow, dose, urn, tow, wart, bruise, caste, ghou, gist, hearse, hearth, pier, plait, raft, rogue, shone, shove, sown, stroll, thine, trough, wan, wand, wharf, yacht, banal, caress, dais, diver, lasso, matrix, parquet, rubric, mousy, palsy, hasty, ably, scary, dial

Pseudoword

List 1. yorse, wough, wouse, wome, wost, woll, lon, han, rone, ase, fruise, sorm, sough, soft, cile, sood, sart, treath, plass, paf, tand, cass, chead, duss, gove, bave, bine, wongy, halger, horry, natric, awal, scully, contrix, diman, danter, dary, moun-try, cosso, poley

List 2. raunt, routh, hild, wamp, heak, rart, woot, wone, mave, arp, squead, slood, swy, fatch, sork, sose, fon, tronce, tarf, colf, tant, pove, groad, breat, dar, gow, colwer, coutle, morine, rubure, hovet, wable, laver, ansine, forand, sponder, crivy, borquet, dion, bary