Consciousness and Cognition 18 (2009) 135–144

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

On the role of set when reading aloud: A dissociation between prelexical and lexical processing

Jeffrey R. Paulitzki [a], Evan F. Risko [b,*], Shannon O'Malley [a], Jennifer A. Stolz [a], Derek Besner [a]

[a] Psychology Department, University of Waterloo, Canada
[b] Psychology Department, University of British Columbia, 2329 West Mall, Vancouver, BC, Canada

ARTICLE INFO

Article history:
Received 10 August 2007
Available online 13 January 2009

Keywords:
Intention
Automaticity
Visual word recognition
Reading

ABSTRACT

Two experiments investigated the role that mental set plays in reading aloud using the task choice procedure developed by Besner and Care [Besner, D., & Care, S. (2003). A paradigm for exploring what the mind does while deciding what it should do. *Canadian Journal of Experimental Psychology, 57*, 311–320]. Subjects were presented with a word, and asked to either read it aloud or decide whether it appeared in upper/lower case. Task information, in the form of a brief auditory cue, appeared 750 ms before the word, or at the same time as the word. Experiment 1 yielded evidence consistent with the claim that at least some pre-lexical processing can be carried out in parallel with decoding the task cue (the 0 SOA condition yielded a smaller contrast effect than the long SOA condition). Experiment 2 provided evidence that such processing is restricted to pre-lexical levels (the word frequency effect was equivalent at the 0 SOA and the long SOA). These data suggest that a task set is a necessary preliminary to *lexical* processing when reading aloud.

1. Introduction

There are large and deep literatures on "attention" (see Pashler, 1998), on visual word recognition, and on reading aloud (see Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Perry, Ziegler, & Zorzi, 2007). However, there has been relatively little work on the intersection of these literatures; that is, the *relation* between "attention" and visual word recognition and reading aloud. In our view this is because the field at large believes the latter processes to be "automatic" and hence does not require any form of attention (e.g., Brown, Gore, & Carr, 2002; LaBerge & Samuels, 1974; Marcel, 1983; Posner & Synder, 1975; see also the many investigators cited in Reynolds & Besner, 2006).

More recent work has yielded a different conclusion in which various forms of attention (mental set, spatial attention, and central attention) are involved in various aspects of visual word recognition and reading aloud (e.g., Besner, Risko, & Sklair, 2005; Bauer & Besner, 1997; McCann, Folk, & Johnston, 1992; O'Malley, Reynolds, Stolz & Besner, 2008; Reynolds & Besner, 2006; Stolz & McCann, 2000; Stolz & Stevanowski, 2004; Risko, Stolz, & Besner, 2005; Lachter, Forster, & Ruthruff, 2004). That said, this work is hardly mainstream; there is little evidence that it has affected cognitive psychology's consciousness in a general way (for example, we have yet to see this view considered at any length in a cognition textbook).

1.1. A question of "set"

The present work concerns itself with the question of "set." For example, we do not typically make posting motions every time we pass a mail-box. Nor do we cross a street just because the light happens to be green. Instead, what we do depends on our plan (mental set). Here we consider the role of mental set in reading aloud.

* Corresponding author.
E-mail address: dbesner@watarts.uwaterloo.ca (E.F. Risko).

1053-8100/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.concog.2008.11.007

The most widely held account of set in the context of visual word recognition and reading aloud (although not explicitly stated this way) assumes that *automatic* processes result in orthographic, phonological, lexical and semantic activation in response to the presentation of a letter string. The *task set* held by the subject determines what to do with the *products* of these automatic processes (e.g., press a key, read aloud). One problem with this idea is that in the experiments used to support it a task set is typically already in place when the subject comes to do the experiment, given that she/he has been instructed what to do, and the task (and hence the task set) is held constant across the time that the subject is engaged with the experiment. This kind of experiment, therefore, does not easily lend itself to investigating the question of whether a task set needs to be in place *before* "stimulus analysis" can begin. A different approach is required in order to assess whether task set *enables* reading aloud. A variant of the task choice paradigm as discussed below offers one way to address this and other questions.

1.2. Set as operationalized in the task-choice paradigm

Imagine that a single word is presented, and skilled readers are asked to read it aloud or signal whether it appears in upper or lower case. The time to do this is measured. On a random half of the trials the word appears in high contrast (bright). On the remaining trials it appears in low contrast (dim). A brief tone 750 ms before the appearance of the word signals the task that the subject has to do on that trial; the task varies randomly from trial to trial. These conditions are so basic that it is hard to imagine anyone taking issue with the claim that words appearing in low contrast would take longer to respond to than words in high contrast.

Suppose, however, that one small change is made. Instead of the task cue always appearing well before the word, on a random half of the trials the onset of the cue appears at the same time as the onset of the word. How should processing of the word be affected by contrast (or other factors) in this particular condition, given the widely held view that processing of a word is initiated by its onset, is ballistic with respect to computing orthographic, phonological and semantic codes, does not need cognitive resources ("attention"), and is not interruptible by other mental activities?

If the subject can decode the task cue at the same time that she/he is encoding the word, and the time to decode the cue is as long (or longer) than the time taken to deal with the effect of contrast (or other manipulated factors such as word frequency) observed when the cue appears well before the word, then the straightforward prediction is that no contrast effect will be observed. This is because the effect of contrast will be completely absorbed into the time taken to decode the task cue. This logic is closely related to the "cognitive slack" logic developed by Pashler that is widely used in the context of the Psychological Refractory Period (PRP) paradigm (see Reynolds & Besner 2006 for an investigation of reading processes in this context).

Besner and Care (2003) reported that when the stimuli consist of pronounceable letter strings that are not words (e.g., "neap"), contrast and SOA had additive effects on the time to read aloud in the context of the task choice paradigm (i.e., there was no evidence for functional letter-string encoding during the time the cue was being decoded). Taken together with other facets of their data, Besner and Care argued that the operations necessary to deal with the effect of contrast in the zero SOA condition did not begin until the subject finished decoding the cue telling her/him which task to perform. They therefore concluded, contrary to the view that dominates the reading literature, that sub-lexical phonological recoding in the service of reading aloud is *not* automatic in the specific sense that it was not initiated by the mere appearance of the letter string, and was not immune to interference from other processes (here, decoding the task cue). In short, Besner and Care concluded that the subject needed a task set to be in place before any functional processing associated with sublexical phonological recoding could take place. In the present investigation, we apply the task choice procedure to the reading of *words* instead of *nonwords*. The experiment described below is the first to explore how words are processed in the context of this paradigm.

1.3. How are words read aloud?

The most successful computational accounts of reading aloud are *dual route* (e.g., Coltheart et al., 2001; Perry et al., 2007). These models generate a phonological code from print by recourse to lexical and non-lexical routines. For present purposes we highlight only the central distinction between these routines.

1.4. The lexical route

Following the activation of feature and letter level processing, orthographic lexical activation occurs, which in turn activates the phonological lexicon, and ultimately speech. Each lexicon (orthographic and phonological) contains a single node (lexical entry) for each unique word the model knows (spelled in the orthographic lexicon; sounded in the phonological lexicon). The lexical route correctly reads aloud all the words the models know. This route is required in order to correctly read aloud exception words (those that violate the typical spelling-sound correspondence rules such as "HAVE" and "PINT").

1.5. The nonlexical route

The nonlexical route translates print into sound sub-lexically via a set of grapheme–phoneme correspondences that have been learned. This route is *required* in order to read non-words aloud (e.g., *neap*) because they have no lexical entries.

The essential point here is that the Besner and Care (2003) result in which stimulus quality was additive with SOA has no necessary implications for what will happen when *words* are read aloud in the context of the task-choice paradigm. In the Besner and Care study all the items were nonwords, and therefore the nonlexical route was required in order to read them aloud. In our first experiment here, all the words were *high frequency*. Such items are typically read aloud by recourse to the lexical route with little if any contribution from the nonlexical route (e.g., see Coltheart et al., 2001). It does not logically follow that the lexical and nonlexical routines always require the same forms of attention. Indeed, although it appears that both words and nonwords require *spatial* attention as a preliminary to further processing (Besner et al., 2005; McCann et al., 1992; Risko et al., 2005; Stolz & McCann, 2000; Stolz & Stevanovski, 2004) this is not the case for the more "central" form of attention indexed by the Psychological Refractory Period (PRP) paradigm. In the context of the PRP paradigm (where a task set is in already in place given that subjects are instructed before the experiment begins as to what they must do to each stimulus), lexical and sublexical processes involved in reading aloud are dissociable. Sublexical phonological processing requires a form of attention, but early nonlexical processing common to both lexical and sublexical processing (features and letters), as well as early *lexical* processing (activation of the orthographic input lexicon), does not (O'Malley, Reynolds, Stolz & Besner, 2008; Reynolds & Besner, 2006). The present experiment thus differs in an important respect from the Besner and Care study despite the use of the same paradigm (task choice), same tasks (reading aloud and case identification), and the same factors (contrast and SOA) in that it addresses the role of task set when reading aloud *words* rather than nonwords.

1.6. Contrast $\times$ SOA: outcomes and accounts

The critical question here thus concerns which outcomes speak to which accounts in the context of the task choice paradigm. Three different outcomes are noted, along with an account for each outcome.

1. If contrast and SOA have additive effects on RT for both tasks then these results parallel those reported by Besner and Care and call for the same interpretation. Namely, a task set is a necessary preliminary to reading aloud words as well as nonwords. It is important to note that the present logic assumes that if the contrast effect is not absorbed into the cognitive slack produced by decoding the task cue (as in Besner & Care, 2003) and thus the process slowed by stimulus contrast is not initiated during cue decoding, no functional "later" process in reading is initiated during cue decoding.
2. A different outcome is that the contrast effect is eliminated at the zero SOA in both tasks. This result would be consistent with the view that task set only comes into play well *after* lexical processing has begun, given that contrast affects lexical processing when reading aloud (low contrast slows low frequency words more than high frequency words when reading aloud, provided that only words appear in the experiment; see O'Malley, Reynolds & Besner, 2007; O'Malley and Besner, 2008; Yap & Balota, 2007).
3. A third outcome is that, at the zero SOA, the contrast effect is reduced but not eliminated (partial underadditivity). This outcome would support the view that *some* processes can proceed without a task set in place, but that others cannot. This outcome would not, on its own, allow the inference that *lexical* processing is possible in the absence of a task set because it is possible that only feature and letter level processing were carried out.

In summary, Experiment 1 examines performance in the task choice paradigm when high frequency words are read aloud given that such items are the most likely to be processed "automatically" if such processing were possible. Thus, the question addressed here concerns whether the task choice paradigm yields results that undermine the hypothesis of "automatic" processing when reading high frequency words aloud or whether a task set is only needed when subjects engage in sub-lexical phonological recoding as required in the Besner and Care (2003) experiment because there subjects were always processing nonwords.

2. Experiment 1

2.1. Method

2.1.1. Subjects

Sixty-four university undergraduates whose first language was English were drawn from the cognition and perception subject pool. Each was paid for his/her participation in a single session lasting 25–30 min.

2.1.2. Stimuli

The stimulus set consisted of 240 monosyllabic, high frequency words (Kucera-Francis frequency between 95 and 6997 occurrences per million) that varied in length from 3 to 5 letters (available from the authors on request). An additional 32 words were generated for use during practice. The stimulus set for the experiment proper was divided into two: half were assigned to the reading aloud task and the remaining half to the case decision task for half the subjects. This assignment of items to tasks was reversed for the remaining subjects. Assignment of items to SOA (0/750), case (lower/upper), cue (low/high tone), and contrast (bright/dim) was also counterbalanced. Stimuli in the bright condition appeared in Micro Experimental Laboratories (MEL) with red, green, blue (RGB) values of 63, 63, 63; stimuli in the dim condition appeared in MEL

Table 1a
Mean RTs (ms) and standard errors (in parentheses) in Experiment 1 as a function of Task, SOA, and Stimulus Quality.

	Reading aloud		Case identification	
	0 SOA	750 SOA	0 SOA	750 SOA
Low stimulus quality	871 (19)	611 (18)	932 (23)	638 (17)
High stimulus quality	825 (14)	540 (12)	930 (25)	614 (22)
Contrast effect	**46**	**71**	**2**	**24**
Size of the interaction (ms)	**25**		**22**	

with RGB values 5, 5, 5. Bright and dim stimuli were intermixed randomly throughout the experiment in a single block of trials.

2.1.3. Placeholders, targets, and cues

On each trial a rectangle appeared in the center of the screen and acted as a focal placeholder that remained on the screen for the remainder of the trial. The rectangle measured 1.5 cm in height by 3.3 cm in width, and was presented in an intermediate grey with corresponding RGB values 29, 29, 29. Inside the rectangle was a second, black rectangle measuring 1.3 cm in height by 3.1 cm in width. The target appeared in white letters on the black background, centered inside the placeholder. A 100 ms duration tone of either low or high pitch (400 or 2550 Hz) served as the task cue.

2.1.4. Apparatus

Subjects viewed targets on a standard 15 in. SVGA color monitor. Vocal responses were recorded by an Optimus 33–1067 microphone triggered by the subject's voice. A standard 104 key Win 95 keyboard was placed within easy reach to allow subjects to execute the button press response required in the case identification task. MEL software running on a Pentium MMX computer was used to control stimulus presentation and record data.

2.2. Procedure

Subjects were tested individually in a dimly lit room. They were seated in front of the computer monitor and received both written and verbal instructions. They were instructed that, on each trial, a grey rectangle would appear on the screen and that they should focus on the inside of the rectangle. Subjects were told that a tone would sound and that a word would appear in the center of the rectangle. Furthermore, they were informed that half of the words would appear in bright form and half would appear dim. Subjects were instructed that, depending on the pitch of the tone, they would either read the word out loud *or* identify the case of the word by pressing the 'z' or 'x' key to indicate lower or upper case, respectively. Subjects used their left hand to make case decisions and dedicated a finger to each of the two response keys.

The 32 practice stimuli were presented in random order for the first block of practice trials. Accuracy was not recorded for this block. A second randomly ordered block of 16 practice trials was then performed repeatedly until an accuracy of 87.5% or higher was attained for the 16 trials. In this fashion, each subject performed at least 48 practice trials before beginning the experiment proper. After the practice phase, subjects were given the opportunity to ask questions and were reminded to respond as quickly and accurately as possible. When ready, subjects were told to press the space bar twice to initiate the experimental trials. Presentation of the target, which appeared in the center of the frame, varied randomly between appearing simultaneously with the tone (0 ms SOA) and being delayed until 750 ms after the onset of the tone presentation (750 ms SOA).

Vocal responses were classified by the experimenter as correct, incorrect or spoiled. An incorrect pronunciation was defined as an addition/deletion or position change of letters. A spoiled trial was one in which the microphone was prematurely triggered by extraneous sounds, when it failed to detect the vocal response, or when the subject executed an inappropriate response (pressing a key) before speaking.[1]

In the case identification task, an appropriate response (pressing of the 'z' or 'x' key) terminated the stimulus display and initiated an inter-trial interval of 1000 ms during which the screen remained blank. In the reading aloud task, a vocal response terminated the display, but the 1000 ms inter-trial interval was initiated by the experimenter's coding of a correct, incorrect, or spoiled response. The MEL software recorded accuracy and RT to the nearest millisecond. All responses were measured from the onset of the target stimulus.

2.3. Results

Incorrect responses (3.1%) and mistrials (3.3%) were discarded. The remaining response time data were subjected to a recursive outlier analysis in which scores falling three or more standard deviations above or below the mean score for each subject in each condition were eliminated from further analyses (Van Selst & Jolicoeur, 1994). This resulted in elimination of 3.2% of the RTs (see Tables 1a and 1b).

[1] In retrospect, we should have tracked inappropriately executed responses (e.g., pressing a key before speaking) separately from spoiled trials in order to analyze the pattern of such task errors.

Table 1b
Mean percentage errors in Experiment 1 as a function of Task, SOA and Stimulus Quality.

	Reading aloud		Case identification	
	0 SOA	750 SOA	0 SOA	750 SOA
Low stimulus quality	2.9	2.1	4.1	2.8
High stimulus quality	0.9	1.4	4.4	2.5
Contrast effect	**2.0**	**0.7**	−**0.3**	**0.3**
Size of the interaction (%E)	−**1.3**		**0.6**	

2.3.1. RTs

The RT data[2] from trials with a correct response were then submitted to an analysis of variance (ANOVA) in which Task (reading aloud vs. case identification), SOA (0 ms vs. 750 ms), and Contrast (bright vs. dim) were within-subjects factors. The main effects were all significant: dim stimuli produced slower RTs than bright stimuli, $F(1,63) = 23.9$, MSe = 6824, $p < .001$, the 0 SOA produced slower RTs than the 750 SOA, $F(1,63) = 959.3$, MSe = 11128, $p < .001$, and case decisions were slower than reading aloud, $F(1,63) = 30.2$, MSe = 18717, $p < .001$.

There were significant two-way interactions between Contrast × Task (larger contrast effect for the reading aloud task), $F(1,63) = 9.2$, MSe = 7139, $p < .01$, and SOA × Task (larger effect of SOA for case identification), $F(1,63) = 8.5$, MSe = 3842, $p < .01$, replicating Besner and Care (2003). Most critically, there was a significant interaction between Contrast and SOA (the effect of contrast was smaller at the short SOA than it was at the long SOA), $F(1,63) = 6.0$, MSe = 3016, $p < .05$. The three-way interaction was not significant ($F < 1$).

2.3.2. Errors

The error data were also submitted to a 3-factor ANOVA in which Contrast, SOA and Task were within-subjects factors. The main effects were all significant: dim stimuli produced more errors than bright stimuli, $F(1,63) = 8.0$, MSe = .0007, $p < .01$, the 0 ms SOA produced more errors than the 750 ms SOA, $F(1,63) = 9.7$, MSe = .0010, $p < .01$, and case decision was more error prone than reading aloud, $F(1,63) = 20.8$, MSe = .0330, $p < .01$. The interaction between Contrast × Task (larger contrast effect for the reading aloud task) was significant, $F(1,63) = 6.2$, MSe = .0052, $p < .05$, as was the SOA × Task interaction (larger effect of SOA for case identification), $F(1,63) = 9.5$, MSe = .0007, $p < .01$. These interactions paralleled the interactions observed in the RT data. No significant interaction was observed between Contrast and SOA ($F < 1$). The 3-way interaction of Contrast × SOA × Task was not significant, $F(1,63) = 3.1$, MSe = .0009, $p > .05$.

2.4. Discussion

The results of the present experiment can be summarized in the following way. First, the contrast effect in the RT data is significantly smaller at the zero SOA than at the long SOA. This supports the inference that at least some processing goes on in parallel with cue decoding (i.e., *some* of the effect of contrast is absorbed into the slack created by the time taken to decode the cue).

The observed underadditivity was complete in the case decision task but not in the reading aloud task (there was no three-way interaction between contrast, task and SOA because the baseline contrast effect was larger in the reading aloud task). As noted in the introduction, partial under-additivity suggests that *some* but not all processes affected by stimulus contrast can occur in parallel with cue decoding. The pattern seen here across tasks suggests that the effect of contrast that is common to *both* case decision and reading aloud can be initiated without a task set in place as evidenced by the complete underadditivity in the case decision task and the partial underadditivity in the reading aloud task. The latter result also suggests that the effect of contrast that is *unique* in the current context of reading aloud (i.e., the part which remains at the zero SOA-the significant 46 ms effect) cannot be initiated without a task set.

It should be acknowledged that the underadditivity observed here between stimulus quality and SOA differs from what Besner and Care observed (additive effects of these same factors). It is not yet clear why this difference is seen. There are differences in the modality of the task cue across experiments. In Besner and Care a visual cue was used whereas here it was auditory. However, we are loath to conclude that this particular methodological difference is responsible for the different outcomes. Indeed, in an experiment using the task choice procedure which also employed an auditory cue, Ansari and Besner (2005) found that when subjects simply named a single Arabic numeral, contrast and SOA had additive effects whereas when subjects had to identify the numeral and add one (e.g., see 8, say 9) there was partial underadditivity of contrast and SOA (indeed, the *reduction* in the size of the stimulus quality effect in that condition was 59 ms, a value twice as large as seen here. A more plausible hypothesis, therefore, is that there are meta set factors in play that are as yet poorly

[2] Given that each trial involved either repeating the same task or doing a new one, an analysis was also performed with trial type (repeat vs. switch trial) as a within-subjects factor resulting in a four-factor ANOVA (Contrast x SOA × Task × Trial Type). For present purposes the important point here is that the four way interaction was not significant in either the RT or error analysis and that none of the three way interactions that included contrast and SOA were significant in either RT or errors (all ps > .12). In other words, the factor of repeating versus switching tasks did not modulate the critical Contrast × SOA interaction that is central here. For purposes of clarity we therefore only report the three factor ANOVAs.

Table 2a
Mean RTs (ms) and standard errors (in parentheses) in Experiment 2 as a function of Task, SOA and Word Frequency.

	Reading aloud		Case identification	
	0 SOA	750 SOA	0 SOA	750 SOA
Low frequency	893 (16)	681 (18)	911 (21)	676 (23)
High frequency	863 (15)	641 (16)	897 (19)	672 (22)
Frequency effect	**30**	**39**	**14**	**4**
Size of the interaction (ms)	**9**		−**10**	

understood. In particular, subjects in the present experiment knew that all the stimuli were words given that only words appeared in the practice trials and the experimental trials were also all words, whereas the subjects in Besner and Care knew that all the stimuli in that experiment were nonwords given that all the practice trials consisted of nonwords and the experimental trials were also all nonwords. To be sure, it will be important for future work to experimentally address why the Besner and Care results differ from the present ones (e.g., one approach would be to randomly mix words and nonwords together). However, for present purposes our interest lies in addressing the role of set when only reading *words* aloud. Consequently we do not consider the discrepancy between the present results and those of Besner and Care any further.

3. Experiment 2

In Experiment 2 subjects were again presented with printed words and either read them aloud or decided whether they appeared in upper or lower case by pressing one of two buttons. The major difference across experiments is that stimulus quality was not manipulated in Experiment 2 (it was in Experiment 1) and word frequency was manipulated in Experiment 2 (it was not in Experiment 1). Given that stimulus quality and SOA yielded an underadditive interaction in Experiment 1, the purpose of Experiment 2 was to determine whether any lexical processing could be carried out during the time that the task cue was being decoded. Lexical processing was indexed by the effect of word frequency. If word frequency and SOA yield an underadditive interaction in the reading aloud task (no effect of word frequency is expected in the case decision task) then this would support the inference that at least some lexical processing can proceed during the time that the task cue is being decoded.

3.1. Method

3.1.1. Subjects

A new set of 64 subjects were recruited from the undergraduate population.

3.1.2. Stimuli

The stimulus set consisted of 160 monosyllabic high frequency words (Kucera–Francis frequency between 150 and 12,144 occurrences per million) and 160 monosyllabic low frequency words (Kucera–Francis frequency between 1 and 75 occurrences per million) that varied in length from 4 to 5 letters (available from the authors on request). In order to ensure that the latter items were processed lexically they were all irregular in terms of their spelling-sound correspondences. According to both the DRC and CDP+ computational models, the correct pronunciation for such items can only be generated by recourse to the lexical route. An additional 32 words (16 high frequency and 16 low frequency) were generated for use during practice. The stimulus set for the experiment proper was divided into two: half were assigned to the reading aloud task and the remaining half to the case decision task for half the subjects. This assignment of items to tasks was reversed for the remaining subjects. Assignment of items to SOA (0/750), case (lower/upper), and cue (low/high tone) was counterbalanced.

3.1.3. Apparatus

E-Prime software was used to control stimulus presentation and record data.

3.2. Procedure

Subjects were given instructions similar to those in Experiment 1, but omitting the instruction that half the words would appear in bright form and half would appear dim. Subjects used the left index finger to press the ‘a’ key and the right index finger to press the ‘l’ key to indicate lower or upper case (counterbalanced across subjects).

3.3. Results

Incorrect responses (2.4%, including incorrect pronunciation and incorrect case decision), trial on which subjects performed the wrong task (2.4%), and mistrials (3.2%) were discarded. The outlier procedure resulted in the exclusion of 1.3% of the data. The remaining data are presented in Tables 2a and 2b.

Table 2b
Mean percentage errors (%E) in Experiment 2 as a function of Task, SOA and Word Frequency.

	Reading aloud		Case identification	
	0 SOA	750 SOA	0 SOA	750 SOA
Low frequency	2.8	3.8	2.9	2.1
High frequency	0.9	1.0	3.5	2.2
Frequency effect	**1.9**	**2.8**	**−0.6**	**−0.1**
Size of the interaction (%E)	**0.9**		**0.5**	

3.3.1. *Task × SOA × Frequency*

The RT data from trials with a correct response were submitted to an ANOVA in which Task (reading aloud vs. case identification), SOA (0 ms vs. 750 ms), and Word Frequency (low vs. high) were within-subjects factors. This analysis yielded significant main effects of frequency in which low frequency stimuli produced slower RTs than high frequency stimuli, $F(1,63) = 28.1$, MSe = 2198, $p < .001$, SOA in which the 0 SOA produced slower RTs than the 750 SOA, $F(1,63) = 839$, MSe = 7634, $p < .001$. There was no main effect of Task, $F(1,63) = 2.1$, MSe = 22310, $p = .148$.

The two-way interaction of Frequency × Task yielded a larger frequency effect for reading aloud than for case decision, $F(1,63) = 10.7$, MSe = 1959, $p < .01$. The SOA × Task interaction was not significant, $F(1,63) = 2.5$, MSe = 2390, $p = .12$. There was no significant interaction between Frequency and SOA, $F(1,63) = .002$, MSe = 2212, $p = .996$. Finally, the 3-way interaction of SOA × Task × Word Frequency was not significant, $F(1,63) = 1.1$, MSe = 2664, $p > .25$.

The error data were also submitted to a 3-factor ANOVA in which Frequency, SOA and Task were within-subjects factors. Low frequency words produced more errors than high frequency words, $F(1,63) = 18.9$, MSe = 6.7, $p < .01$. However, there were no significant effects of SOA, $F(1,63) = 1.1$, MSe = 6.5, $p > .1$, or Task, $F(1,63) = 2.1$, MSe = 20.4, $p > .1$. There was a larger frequency effect for the reading aloud task, than for the case decision task, $F(1,63) = 25.3$, MSe = 8.9, $p < .01$, a larger effect of SOA for case identification than for reading aloud, $F(1,63) = 17.7$, MSe = 4.7, $p < .01$, and a marginally smaller frequency effect at the 0 ms SOA than at the long SOA, ($F(1,63) = 3.7$, MSe = 4.3, $p = .06$). The 3-way interaction of Frequency × SOA × Task was not significant ($F < 1$).

3.3.2. *Frequency × SOA in reading aloud*

Given that word frequency should affect the reading aloud task, but there is no a priori reason it should also affect the case decision task, we also analyzed the reading aloud task on its own. In the 2-way ANOVA of Word Frequency × SOA on the RT data the main effects of both Word Frequency and SOA were significant, $F(1,63) = 41.5$, MSe = 1863, $p < .01$, and $F(1,63) = 587.6$, MSe = 5121, $p < .01$ respectively. The interaction between Frequency × SOA was not significant ($F < 1$).[3]

For the errors, a two-way ANOVA reflected the same pattern observed in the RT data; there were significant main effects of Frequency, $F(1,63) = 39.2$, MSe = 8.8, $p < .01$, and SOA, $F(1,63) = 4.5$, MSe = 4.6, $p < .01$, but no significant interaction between Frequency and SOA, $F(1,63) = 2.3$, MSe = 5.6, $p = .13$.

3.4. *Effects of task repetition*

On some of the trials subjects do the same task as on the previous trial (a "stay" trial), and on other trials they do the other task (a "switch" trial). An obvious question questions concerns whether the pattern of results is different for switch trials versus stay trials (e.g., see Besner & Risko, 2005). We therefore conducted a subsidiary analysis on the reading aloud data which coded for task type on the previous trial. If the previous trial also consisted of reading aloud then the current trial would seem to be the most likely to pick up effects of encoding the word lexically during the time taken to decode the cue given that the correct set is already in place. This analysis yielded significant main effects of SOA, $F(1,63) = 585.6$, MSe = 10264, $p < .01$, Word Frequency $F(1,63) = 47.9$, MSe = 3503, $p < .01$, and of last task $F(1,63) = 22.8$, MSe = 5327, $p < .01$, where if the last task was the same as the current task (reading aloud) participants were faster than if the previous task was different (case decision). The only 2-way interaction that was significant was SOA × last task, in which "stay" trials yielded a smaller SOA effect than on switch trials. No other effects were significant (Fs < 1). These results show that there is a repetition benefit on stay trials, evidenced by the fact that responses were faster, and that the size of the SOA effect was smaller on stay trials. However, this did not modify the size of the word frequency effect across SOA (indeed, the interaction of word frequency × SOA was (non-significantly) 4 ms *smaller* on stay trials rather than being larger than when the task on the previous trial was different (see Table 3a).

Finally, for the errors, a 3-way analysis looking at last task, SOA and frequency in the reading aloud task, there was a main effect of frequency, $F(1,63) = 40.4$, MSe = 17.5, $p < .01$, and a main effect of SOA, $F(1,63) = 4.9$, MSe = 9.6, $p < .05$. No other effects were significant (Fs < 2) (see Table 3b).

[3] We also computed medians to get another estimate of the size of the interaction between word frequency and SOA in the reading aloud task. This analysis yielded an interaction estimate of 3 ms ($F < 1$).

Table 3a
Mean RTs (ms), in the reading aloud task, Experiment 2, as a function of the task completed on the previous trial, SOA and Word Frequency.

Previous task	Reading aloud		Case identification	
	0 SOA	750 SOA	0 SOA	750 SOA
Low frequency	869	677	919	685
High frequency	833	636	889	645
Frequency effect	**36**	**41**	**30**	**40**
Size of the interaction (ms)	**5**		**10**	

Table 3b
Mean percentage errors in the reading aloud task from Experiment 2 as a function of the task completed on the previous trial, SOA and Word Frequency.

Previous task	Reading aloud		Case identification	
	0 SOA	750 SOA	0 SOA	750 SOA
Low frequency	2.5	3.6	3.1	4.0
High frequency	.6	1.3	1.1	.8
Frequency effect	**1.9**	**2.3**	**2.0**	**3.2**
Size of the interaction (%E)	**0.4**		**1.2**	

3.5. Discussion

The data from Experiment 2 are clear in that there is no support for the idea that *lexical* level processing goes on in parallel with cue decoding inasmuch as the interaction between word frequency and SOA was not significant ($F < 1$). A further analysis which considered only the reading aloud task as a function of the task that took place on the prior trial produced the same result: no under-additive interaction between word frequency and SOA ($F < 1$).

3.6. Considerations of power

We also conducted a power analysis on Experiment 2, using the effect size observed in Experiment 1 for the reading aloud task. There was .85 power to detect a 25 ms under-additive interaction between word frequency and SOA. Given that power drops off considerably as the effect size gets smaller, we can not rule out the possibility that there is a small under-additive interaction that we can not detect with this relatively large sample size. On the other hand, Experiment 1 yielded a substantial (46 ms) stimulus quality effect in the reading aloud task that was *not* absorbed into cognitive slack. We are therefore inclined to the view that there is little in the way of a genuine under-additive interaction between lexical processing and SOA.

4. General discussion

The present experiments were conducted in order to determine whether mental set, operationalized as having a task set in place prior to processing the target, plays any role in reading *words* aloud. The standard account is that mental set plays a role only *after* automatic processes have identified the word, and generated a phonological code. The results of the present experiments qualify this widely received theoretical view.

At the empirical level the results from Experiment 1 are novel in that this is the first report of both tasks yielding a reduction in the magnitude of the contrast effect at the zero SOA when reading *words* aloud. This result suggests two major conclusions. First, *early* processes (e.g., feature and letter processing), at least when only words appear in the experiment, can proceed during the time taken to decode the task cue. In this context then, task set is *not* a necessary preliminary for target processing to begin. This result is thus in accord with the prevailing "automaticity" view.

The results of Experiment 2, also qualify the conclusion regarding automaticity because they suggest that task set does play a rather more important role than generally acknowledged when it comes to *lexical* processing of print. Rather than being subordinate to "automatic" processing triggered by the onset of a stimulus, functional *lexical* processing, as indexed here by the effect of word frequency, depends on first implementing a task set that enables these processing operations. That is, the failure to see an under-additive interaction between word frequency and SOA in Experiment 2 suggests that functional lexical processing awaits the interpretation of the cue so that subjects "know" what task they have to carry out. To our knowledge, this is the first result with intact subjects that questions the wide spread idea that ***lexical*** processing when reading words aloud is not affected by mental set in the form of a task set.

The results from Experiment 2 also contrast with those reported in the literature on the PRP. In this paradigm subjects are typically presented with a stimulus (S1) and asked to categorize it in some way (e.g., a tone is presented and it is categorized as high or low pitch with an overt response). A second stimulus is also presented (S2) and subjects also categorize that stimulus (e.g., a letter string is presented and subjects make a lexical decision or read it aloud). Two factors are typically manipulated. One is the difficulty of processing S2 and the other is the SOA between S1 and S2. The standard result is that the time to respond to S2 gets slower as the SOA between S1 and S2 *decreases* (not surprising given that subjects are instructed to respond to S1 before they respond to S2). A major question concerns whether it is possible to carry out any mental work on S2 while mental work is being carried out on S1. If such work is possible, then the effect of the difficulty manipulation on S2 should get *smaller* as SOA decreases because work that is carried out on S2 in parallel with work on S1 will be absorbed into the slack created by having to respond to S1 before S2 (see Pashler, 1984).

For present purposes one of the critical differences between the task choice paradigm used in the experiments reported here and the PRP paradigm is that in the latter paradigm there is no uncertainty about the task that needs to be performed on S2. That is, it is fixed, and hence is always known before the trial starts. In short, a mental set is always already in place in the PRP paradigm as described here, but in the task choice paradigm the subject does not know what task to perform until the cue has been decoded. Without claiming that this difference is an exhaustive description of all differences between these paradigms, this distinction does permit an inference of interest to those with reservations as to whether the results of the present Experiment 2 are evidence that lexical processing needs a task set in place in order to begin. To put the issue another way, if lexical processing of S2 in the PRP paradigm is at least partially under-additive with decreasing SOA *where a task set is already in place*, then the present observation of a failure to see such under-additivity between lexical processing and decreasing SOA in the context of the task choice paradigm is consistent with the argument that a task set is needed to allow lexical processing to begin. Without claiming that under-additivity of lexical processing in the context of the PRP paradigm is always seen, is complete, and that there are no complicating factors, it is clear that there are numerous reports of such under-additivity (e.g., Allen et al., 2002; Clelland, Gaskell, Quinlan & Tamminen, 2006; Reynolds & Besner, 2006; O'Malley, Reynolds, Stolz & Besner, 2008; Rabovsky, Alvarez, Hohlfeld, & Sommer, 2008; Ruthruff, Allen, Lien, & Grabbe, 2008).

4.1. Is there enough slack?

In Experiment 1 there was a substantial sized stimulus quality effect in the reading aloud task that was not absorbed into slack (46 ms). One potential explanation for this partial under-additivity is that there was not enough "slack" to absorb the entire stimulus quality effect. If this is the case then the lack of any under-additivity in Experiment 2 might be due to there not being enough slack to absorb the word frequency effect, provided the stage of processing influenced by word frequency occurs after that influenced by stimulus contrast.

Previous research using the task choice paradigm suggests that there is more slack available than is being "used" in the present experiments. For example, using the task choice paradigm Ansari and Besner (2005) reported a 59 ms of absorption of a stimulus quality effect. Critically, the amount of slack present in that experiment was comparable to the amount of slack present in the current experiments (using the SOA effect as a rough index). Thus, we would argue that it is the nature of the task (i.e., reading aloud) and its reliance on a task set being in place that determines the amount of absorption rather than the amount of slack available. It remains to be explained, therefore, why all of the stimulus quality effect in Experiment 1 was not absorbed into slack. It may be that there is a currently unidentified process that is task set dependent intervening between feature analysis and the *products* of letter analysis (conversion into abstract letter identities?).

The issue of how much cognitive slack is available to absorb the effects of a given factor has attracted very little consideration in studies using the task choice paradigm and even the more popular PRP paradigm (where cognitive slack logic was developed). In the task-choice paradigm, the amount of cognitive slack at the 0 ms SOA could be considered to be equivalent to the amount of time required to encode and decode the auditory cue and to implement the task set signaled by the cue. It is not until this series of processes is completed that there is a task set in place. Thus, we would be very surprised if the duration of slack available was substantially less than between 100 and 200 ms. Indeed, using the typical estimate of cognitive slack, the SOA effect, the amount of cognitive slack in Experiments 1 and 2 would be approximately 200–300 ms.

If we assume that the amount of cognitive slack is approximately 200 ms, we can then use estimates of when the effects of stimulus contrast and word frequency occur to determine if it is reasonable to expect them to be absorbed into this amount of slack. Using an ERP, measure, stimulus contrast has been demonstrated to influence processing as soon as 80–110 ms after stimulus presentation (e.g., Johannes, Munte, Heinze, & Mangun, 1995). In addition, Sereno and Rayner (2003), using behavioral, eye tracking, and ERP indices, estimated lexical access (a stage widely assumed to be sensitive to word frequency) to occur approximately 150 ms after stimulus presentation. Thus, both stimulus contrast and word frequency influence processes that occur within 200 ms of stimulus presentation and therefore should (if our estimate of slack is correct) be absorbed into the cognitive slack introduced by task cue processing. Of course, this argument fails if our estimate of the amount of cognitive slack is incorrect (but if it fails then an explanation would be needed for the effect of SOA—particularly so when the task repeats from trial N to trail $N + 1$). Future work manipulating the amount of slack available (e.g., by increasing the length of time it takes to encode, decode, and implement a task set) will be useful in determining how much (if any) absorption we can expect if a process does not require a task set.

5. Conclusion

It is a received idea that skilled readers "read" automatically, and one definition of such automaticity is that it occurs in the absence of a task set. We take the present data (the persisting stimulus quality effect in Experiment 1 at the zero SOA, and the full blown word frequency effect at the zero SOA in Experiment 2) to suggest that this idea needs qualification. Some (but not all) sub-processes involved in reading aloud need a task set in place before it is possible for these processes to operate.

Acknowledgments

This work was supported by a summer undergraduate research fellowship from the Natural Sciences and Engineering Research Council of Canada (NSERC) to JRP, a NSERC Canada graduate scholarship and a Killam post-doctorate fellowship to EFR, and NSERC Discovery Grants 183905 and AO998 to JAS and DB, respectively.

References

Allen, P. A., Lien, M. C., Murphy, M. D., Sanders, R. E., Judge, K. S., & McCann, R. S. (2002). Age differences in overlapping-task performance. Evidence for efficient parallel processing in older adults. *Psychology and Aging, 17*, 505–519.

Ansari, I., & Besner, D. (2005). A role for set when naming Arabic numerals: How intentionality limits (putatively automatic) performance. *Psychonomic Bulletin & Review, 12*, 1076–1081.

Bauer, B., & Besner, D. (1997). Processing in the Stroop task: Mental set as a determinant of performance. *Canadian Journal of Experimental Psychology, 51*, 61–68.

Besner, D., & Care, S. (2003). A paradigm for exploring what the mind does while deciding what it should do. *Canadian Journal of Experimental Psychology, 57*, 311–320.

Besner, D., & Risko, E. F. (2005). Stimulus-response compatible orienting and the effect of an action not taken: Perception delayed is automaticity denied. *Psychonomic Bulletin and Review, 12*, 271–275.

Besner, D., Risko, E. F., & Sklair, N. (2005). Spatial attention as necessary preliminary to early processes in reading. *Canadian Journal of Experimental Psychology, 59*, 99–108.

Brown, T. L., Gore, C. L., & Carr, T. H. (2002). Visual attention and word recognition in Stroop color naming: Is word recognition automatic? *Journal of Experimental Psychology: General, 131*, 220–240.

Cleland, A. A., Gaskell, M. G., Quinlan, P. T., & Tamminen, J. (2006). Frequency effects in spoken and visual word recognition: Evidence from dual-task methodologies. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 104–119.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256.

Johannes, S., Munte, T. F., Heinze, H. J., & Mangun, G. R. (1995). Luminance and spatial attention effects on early visual processing. *Cognitive Brain Research, 2*, 189–205.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293–323.

Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after Broadbent: Still no identification without attention. *Psychological Review, 111*, 880–913.

Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology, 15*, 197–237.

McCann, R. S., Folk, C. L., & Johnston, J. C. (1992). The role of spatial attention in visual word processing. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 1015–1029.

O'Malley, S., & Besner, D. (2008). Reading aloud: Qualitative differences in the relation between stimulus quality and word frequency as a function of context. *Journal of Experimental Psychology: Learning Memory & Cognition, 34*, 1400–1411.

O'Malley, S., Reynolds, M. G., & Besner, D. (2007). Qualitative differences between the joint effects of stimulus quality, word frequency in reading aloud, lexical decision: Extensions to Yap and Balota (2007). *Journal of Experimental Psychology: Learning, Memory And Cognition, 33*, 451–458.

O'Malley, S., Reynolds, M. G., Stolz, J. A., & Besner, D. (2008). Reading aloud: Spelling-sound translation uses central attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 422–429.

Pashler, H. (1998). *The psychology of attention*. Cambridge, MA: The MIT Press.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review, 114*, 273–315.

Posner, M. I., & Synder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The loyola symposium* (pp. 55–85). Hillsdale, HJ: Erlbaum.

Rabovsky, M., Alvarez, C. J., Hohlfeld, A., & Sommer, W. (2008). Is lexical access autonomous? Evidence from combining overlapping tasks with recording event-related potentials. *Brain Research, 1222*, 156–165.

Risko, E. F., Stolz, J. A., & Besner, D. (2005). Basic processes in reading: Is visual word recognition obligatory? *Psychonomic Bulletin & Review, 12*, 119–124.

Reynolds, M., & Besner, D. (2006). Reading aloud is not automatic: Processing capacity is required to generate a phonological code from print. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1303–1323.

Ruthruff, E., Allen, P. A., Lien, M. C., & Grabbe, J. (2008). Visual word recognition without central attention: Evidence for greater automaticity with greater reading ability. *Psychonomic Bulletin & Review, 15*, 337–343.

Stolz, J. A., & McCann, R. S. (2000). Visual word recognition: Reattending to the role of spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1320–1331.

Stolz, J. A., & Stevanovski, B. (2004). Interactive activation in visual word recognition: Constraints imposed by the joint effects of spatial attention and semantics. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 1064–1076.

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cognitive Science, 7*, 489–493.

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology, 47A*, 631–650.

Yap, M. J., & Balota, D. A. (2007). Additive and interactive effects on response time distributions in visual word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 33*, 274–296.