A change of task prolongs early processes:

evidence from ERPs in lexical tasks.

Heike Elchlepp, Aureliu Lavric and Stephen Monsell

University of Exeter

Correspondence concerning this article should be addressed to S.Monsell@exeter.ac.uk

Dr. Stephen Monsell
Psychology, College of Life and Environmental Sciences
Washington Singer Labs
University of Exeter
Exeter EX4 4QG
UK

ABSTRACT

Switching tasks costs time. Allowing time to prepare reduces the cost, but usually leaves an irreducible "residual cost". Most accounts of this residual cost locate it within the response-selection stage of processing. To determine which processing stage is affected, we measured event-related potentials (ERPs) as participants performed a reading or a perceptual judgment task and examined the effect of a task-switch on early markers of lexical processing. A task cue preceding a string of blue and red letters instructed the participant either to read the letter string (for a semantic classification in Experiment 1, a lexical decision in Experiment 2), or to judge the symmetry of its color pattern. In Experiment 1, having to switch to the reading task delayed the evolution of the effect of word frequency on the reading task ERP by a substantial fraction of the effect on RT. In Experiment 2, a task switch delayed the onset of the effect of lexical status on the ERP to about the same extent that it prolonged the RT. These effects indicate an early locus of (most of) the residual switch cost: we propose that this reflects a form of task-related *attentional* inertia. Other findings have implications for the automaticity of lexical access: effects of frequency, lexicality and orthographic familiarity on ERPs in the symmetry task indicated involuntary but attenuated orthographic and lexical processing even when attention was focused on a non-lexical property.

It is generally harder to switch among several cognitive tasks than to continue to perform just one. A popular way to capture at least part of this multitasking overhead in the laboratory is to conduct a reaction time (RT) experiment in which, on each of a series of trials, a stimulus is presented, and the participant must perform one of several possible tasks. On some trials the task changes, on others it does not, other things being equal. Several such "task-switching" paradigms have been developed and extensively explored over the last two decades. (For reviews see Kiesel, Steinhauser, Wendt, Falkenstein, Jost, Philipp, & Koch, 2010; Monsell, 2003; Vandierendonck, Liefooghe, & Verbruggen, 2010). The most obvious phenomenon they reveal is the substantial cost to performance of a switch of task: the average RT for a correct response is longer (and there are usually more errors) when the task changes than when it does not.

The opportunity to prepare for a task before presentation of the stimulus generally reduces this *switch cost*, but does not eliminate it (Meiran, 1996; Rogers & Monsell, 1995). To manipulate preparation time unconfounded by the time elapsing since the previous performance of a task, *task-cuing* can be used (Meiran, 1996): each stimulus follows the previous response by a standard interval, and a task cue precedes the stimulus by an interval which is independently manipulated to vary preparation time. As the cue-stimulus interval (CSI) is increased from zero, the RT switch cost reduces to an asymptotic *residual cost*, usually reached at a CSI of between half a second and one second (see Monsell & Mizon, 2006, for examples). This suggests the switch cost has two components: one that can be mitigated by the exercise of active voluntary control (reflected in the reduction in switch cost, if there is time for preparation), and one that cannot — the *residual cost*. We focus here primarily on this residual cost. As we review in more detail below, most accounts of the residual cost assume that what causes the extra difficulty on a switch trial, even after active preparation, is the application of the required set of stimulus-response rules in the face of the passive persistence (from the previous trial, or from traces in

long-term memory) of the rules associated with the competing task, and/or of their involuntary reactivation (by the stimulus). This implies that the residual cost has its effect primarily on the "response-selection" stage of processing.

In the experiments reported in this paper, we used event-related potentials (ERPs) as a chronometric tool (Meyer, Osman, Irvin, & Yantis, 1988) to test this assumption about the locus of the effect, by determining how early within the latent interval (between stimulus and response) processing is prolonged by a prepared switch. Attributing the residual switch cost to an increase in the duration of the response-selection stage implies a relatively late locus of the effect. To partition early and late processing stages we had participants switch between a task requiring lexical processing and a non-lexical task. We measured how a task-switch modulated an ERP marker of lexical processing: the effect of frequency in Experiment 1, the effect of lexical status (word versus non-word) in Experiment 2. To anticipate, the results of both experiments indicate that task switching delayed lexical access by a substantial fraction of the total RT switch cost, suggesting an early locus of (a substantial part of) the residual cost.

Because we had participants switch between a task requiring lexical access and a task requiring attention to non-linguistic properties – the distribution of colors across the letter string – the experiments are also informative concerning another controversial issue: the "automaticity" of lexical access. It used to be a commonplace claim that after extensive practice at reading, recognition of an attentively fixated familiar word becomes "automatic" (e.g., Brown, Gore, & Carr, 2002; LaBerge & Samuels, 1974, Posner & Snyder, 1975) - implying involuntariness and lack of attentional resource demand, or both. The Stroop effect — the marked interference observed when a color name is printed in a conflicting color and participants must name the font color (Stroop, 1935) — has in particular been interpreted as showing that, for skilled readers, recognition of a familiar word and access to its meaning or phonology is "automatic", in the

sense of involuntary. But limits to the involuntariness of lexical access, as indexed by the Stroop effect, have also been demonstrated, as we review below. In the present experiments, we could assess the extent to which ERP markers of lexical access seen when the participant performed the lexical task also appeared when they performed the other, non-linguistic, task.

Theories of the (residual) task-switch cost, and its locus

In most task-switching experiments, two or more sets of S-R rules are defined so that for the *incongruent* stimuli — those to which different responses must be made in the two tasks — the appropriate response would be ambiguous without the task cue. Stimuli mapped to the same response in both tasks are said to be *congruent,* stimuli affording a response in only one task are *univalent*. Competition between task-sets is manifest at two levels. Competition at the S-R level is indicated by a response congruence effect: responses are slower and less accurate to incongruent than to congruent stimuli (Brown, Reynolds, & Braver, 2007; Goschke, 2000; Meiran, 2000; Meiran & Kessler, 2008; Rogers & Monsell, 1995; Wendt & Kiesel, 2008). Competition at the task-set level is indicated by longer RTs to congruent stimuli — for which mere S-R activation via the irrelevant task-set should facilitate response selection, than to univalent stimuli — which should activate neither the irrelevant task-set nor its responses (Aron, Monsell, Sahakian, & Robbins, 2004; Rogers & Monsell, 1995; Steinhauser & Hübner, 2007), and by other evidence (Elchlepp, Lavric, Mizon, & Monsell, 2012; Monsell, Taylor, & Murphy, 2001).

Many theorists (though not all – see Logan & Bundesen, 2003; Schneider & Logan, 2005[1]) assume that in the task-cuing paradigm, the participant interprets the cue and actively engages the appropriate task-set. When the task changes, this requires a process of "task-set

---

[1] These authors have argued that in the task-cuing paradigm, participants retrieve the response specified by the compound of stimulus and cue: there is no change of task-set. For arguments against this account, see Forrest, Monsell, & McLaren (2014).

reconfiguration" (TSR), which may include re-allocating attention and/or activating the appropriate set of S-R rules and/or suppressing the previously active set (Rogers & Monsell, 1995). If the task is the same as on the previous trial, participants might or might not actively reinforce the existing task-set via the same process, but this should not involve such a substantial modulation of task-set parameters, if any; we might say it does not constitute *re*-configuration. The reduction in switch cost with preparation suggests that a motivated participant can, if time permits, carry out some part of TSR before the stimulus arrives. Doing so reduces RT because it (a) saves time that would otherwise be consumed by TSR after the stimulus onset, and/or (b) reduces competition from the irrelevant task-set (response congruence effects sometimes do, and sometimes do not reduce with preparation; see Monsell & Mizon, 2006), and/or (c) avoids the dual-process cost of concurrently performing TSR processes and the task-specific processes required.

Clearly, whatever is accomplished by task-set reconfiguration on a switch trial does not completely disable the other task-set(s). On task-repeat trials, competition from the irrelevant task-set is still evident in S-R level and task-set level interference, although such interference is usually smaller than on a task-switch trial (e.g., Goschke, 2000; Rogers & Monsell, 1995). A substantial congruence effect can still be seen several trials after a switch when the transient after-effects of a task switch are no longer detectable in overall performance (Monsell, Sumner, & Waters, 2003). Indeed, it may be non-optimal to suppress the irrelevant task-set completely, as it must be kept available for reactivation when the task switches again. The need to maintain multiple task-sets in an available state provides a possible reason for the "mixing cost" – the longer RT on task-repeat trials in a mixed-task block than in a pure-task block (Braver, Reynolds, & Donaldson, 2003; Fagot, 1994, reported in Pashler, 2000).

The residual switch cost clearly indicates that advance preparation on a task-switch trial does not achieve the degree of task-readiness manifest on task-repeat trials. The literature offers a number of theories about why this is so, which we now summarize in turn, focusing on what each theory has to say about which processing stage is prolonged on a switch trial:

Exogenously triggered completion of reconfiguration. Rogers and Monsell (1995), having attributed the reduction in switch cost with preparation to a TSR process, suggested the residual cost might be due to a second stage of TSR dependent upon presentation of the stimulus. Rubinstein, Meyer and Evans (2001) speculated that the task-goal could be retrieved prior to the stimulus, but retrieval of the S-R rules can only happen after stimulus identification. Any such post-stimulus component of TSR would need to be largely completed before response selection begins, so that the S-R rules appropriate to the current task have become more active than the alternative set(s). Hence this class of theory implies a relatively early locus of the residual cost.

Task-set inertia. Allport, Styles, and Hsieh (1994) suggested that the task-switch cost resulted from "a form of pro-active interference from features of the task set implemented in previous tasks" (p.421). They labeled this persistence of a previous task-set "task-set inertia", and proposed that the resulting conflict with the now-appropriate task-set results in "additional time for the system to settle to a unique response decision (or response retrieval) after the next imperative stimulus has arrived" (p.437). When it was found (Meiran, 1996; Rogers & Monsell, 1995) that the (mean) switch cost seemed to have two components — one eliminable by active preparation, one not — the idea of passive "inertia" seemed appropriate to account for the latter — the residual cost. Allport et al. (1994) initially conceived of task-set inertia as having a persistence of the order of minutes, but such long-term effects of prior encounters (Allport & Wylie, 2000) seem better dealt with by associative activation of task-sets by stimuli, as later proposed by Waszak, Hommel and Allport (2003) and discussed below.

Task-set inertia could incorporate either persisting activation of the previous task-set or persisting inhibition of the now-appropriate task, or both (Allport et al., 1994; Masson, Bub, Woodward, & Chan, 2003; Yeung & Monsell, 2003; see Vandierendonck et al., 2010, for review). Hence, the core idea of task-set inertia is that persisting activation of task-set activation and/or inhibition from the most recent trial, or last few trials, is a major source of conflict and the (residual) switch cost (Altmann & Gray, 2008; Gilbert & Shallice, 2002; Mayr & Keele, 2000; Meiran, Chorev, & Sapir, 2000; Yeung & Monsell, 2003). For present purposes, the critical question is which stage of processing might be prolonged by task-set inertia. In principle there could be "inertia" in any component of a task-set, but both Allport et al. (1994) and Yeung and Monsell (2003) assumed that the locus of the conflict produced by task-set inertia was in response selection. Other loci are certainly plausible. Masson et al. (2003) observed greater difficulty switching from color naming to word naming when the stimulus whose color was named was a word rather than a string of asterisks and argued that this reflected persisting inhibition of the word-reading pathway. Yeung, Nystrom, Aronson and Cohen (2006) conducted an fMRI study in which participants switched between short blocks of classifying the face or the word in a composite of the two. In a separate part of the experiment, they located (in the same participants) regions of the right fusiform gyrus and left inferior temporal gyrus selectively activated by face and word processing, respectively. In the switching blocks the degree of activation of a task-related region predicted the RT cost of switching *away* from that task. As well as providing a neural correlate of task-set inertia, this is more suggestive of persisting activation of a whole domain-specific processing pathway than of S-R rules per se.

Response-contingent reconfiguration. While preparation helps, the residual cost implies that task-set reconfiguration is incomplete until the new task has been performed at least once. Meiran et al. (2000) argued that while attention could be redirected to a different dimension

during the preparation interval, removing the cost to processing of inappropriate attentional orientation, it is only when the participant actually responds that response-set biasing (changing the meaning of the response set) can be accomplished – hence the residual cost. But more recently, Meiran and colleagues concluded on the basis of simulations of the computational model CARIS (Meiran, Kessler, & Adi-Japha, 2008) that response-set biasing may well occur before stimulus onset, whereas attentional re-setting occurs mainly/only after the stimulus has been presented.

A related claim about the importance of making a response was made by Schuch and Koch (2003) on the basis of an observation since replicated and extended in a number of studies (see Los & van der Burg, 2010, for a review). On some trials they presented a 'no-go' signal along with the target stimulus, telling the subject to withhold a response. This eliminated the switch cost on the following trial. They argued that "when switching from one task to another in typical task-switching paradigms, a conflict arises between the previous and the current category–response rules, and it becomes necessary to inhibit the previous set of rules" (p.101). That is, even after full pre-stimulus preparation, conflict arises after the stimulus onset, and a task-set suppression process must be applied to resolve it and ensure that the appropriate response is produced. Until that suppression has been applied, task-set reconfiguration is incomplete and does not "stick". (According to Koch and colleagues, this suppression process is also the source of the n-2 repetition effect - see Koch, Gade, Schuch, & Philipp, 2010, for review). This is essentially an inhibitory variant of the task-set inertia hypothesis (i.e. inhibition of a task-set on the previous trial persists) coupled with the assumptions that effective inhibition can be applied only post-stimulus (not during preparation), specifically during response selection, not before. The locus of the extra processing time induced by task-set inertia is thus the response selection process.

Associative reactivation of a competing task-set. Waszak et al. (2003) proposed that switch costs arise not because task-set activation carries over from the previous trial, but because the stimulus reactivates (retrieves) task-sets with which it was previously associated. They found slower responses to stimuli presented earlier in the context of the competing task. An attractive feature of this theory is that it provides a rationale for why there are long-term effects: they arise from associative learning, rather than transient persistence of activation. It also potentially offers an account of why advance preparation is only partially effective: as in the Schuch and Koch account, it may not be possible to suppress the source of conflict until conflict happens, and it is only after the stimulus has reactivated the competing task-set that conflict occurs. However, it should be noted that for the theory to provide an account of the residual switch cost, associative activation must have a greater effect on switch than on repeat trials. Thus, while the theory was originally proposed by Waszak et al. as an alternative to task-set inertia, Waszak, Hommel and Allport (2005) acknowledged that task-set inertia is also required to explain why switch trials are particularly vulnerable to the effects of task-set reactivation. This interaction between associative history and switch/repeat remains controversial (e.g. Koch & Allport, 2006). But other evidence for a "retrieval" rather than "persistence" account of task-set interference has recently been advanced by Mayr and colleagues (Bryck & Mayr, 2008; Mayr, Kuhns, & Hubbard, 2014). They examined a phenomenon widely seen as a reflection of task-set inertia – a greater cost of switching to the task that is "stronger" (more practiced/familiar/habitual) than to the "weaker" task. Mayr and colleagues demonstrated that a similar asymmetry in the performance cost was caused by a brief interruption or pause in a sequence of trials of just one task, but only when the stimuli had earlier occurred in the context of an alternative task. Because that alternative task had not been performed for some time, the observed asymmetry could not be

due to transient trial-to-trial "persistence" but must reflect some longer-term association between stimulus and task.

Which processing stage might be slowed by reactivation of a competing task set? Any component of a task-set – attentional settings, S-R rules, etc — could in principle be reactivated by the stimulus. But, as for task-set inertia, the default assumption seems to have been that response-selection is the primary locus.

Probabilistic failure to engage the task-set. De Jong (2000) proposed that both the reduction in switch cost with increasing the preparation interval and the residual cost were explained by just one all-or-none TSR process that sometimes "fails to engage". In particular, he speculated that what fails to engage is the "associative encoding of a cue-action pair" (p.364), in other words the S-R mapping. Hence preparation was either completely effective (performance as good as on a task-repeat trial) or completely ineffective (performance as bad as on an unprepared switch trial). The residual switch cost arises from the average of these two outcomes weighted by their probability. The main evidence favoring this account is that (in the alternating runs paradigm, where task-repeat trials show no effect of preparation) RT distributions on prepared switch trials could be modeled as a mixture of RTs from unprepared switch and prepared repeat trials (De Jong, 2000; Nieuwenhuis & Monsell, 2002; but see Lien, Ruthruff, Remington, & Johnston, 2005, for an argument against this account). The main puzzle is that long preparation intervals, which should allow multiple attempts to engage the task-set, still leave a residual cost in highly motivated participants. For present purposes, De Jong's account is a late — response-selection — locus of the residual cost, although in principle any component of task-set could fail to engage.

Thus far, we have reviewed a number of accounts of the residual cost of a task-switch, of which the most popular is some form of task-set inertia, whether persisting activation or

persisting inhibition, or both, possibly coupled with associative reactivation of the task-set (Waszak et al., 2003). While these accounts do not all logically require that the locus of the residual cost is response-selection, that seems to have been the default assumption (e.g., Allport et al., 1994; Meiran, 2000; Waszak et al, 2003; Yeung & Monsell, 2003), and a late locus appears to be intrinsic to the Schuch and Koch (2003) account. We used event-related potentials to localize the effect of a task-switch within the latent interval. To anticipate, we found evidence for a relatively early locus for at least a substantial portion of the switch cost. One interpretation of this we will consider is that there is substantial inertia in attentional settings from trial to trial, and we will review in the Discussion some other evidence for attentional inertia.

The experiments

In principle, event-related potentials provide a chronometric tool for partitioning the reaction time into the durations of underlying processing stages (Meyer et al., 1998), and allocating effects of an experimental variable on RT, such as task-switch versus -repeat, amongst them. To do this, one must identify ERP components, or markers, detectable in the ERPs, that can be confidently tied to processes of interest. Difficulties include the interpretation of potential markers, the variability of process durations, the possibility of overlapping process durations, and the need to average over many trials to extract a clear ERP signal from the noise. Hence, such chronometric applications of ERP remain relatively rare. However, given a suitable marker, one should at least be able to distinguish between the prolongation of "early" and "late" processes by a task-switch.

To engage in this chronometric enterprise one needs a task for which there is some agreement on the constituent processing stages, a variable known to affect one such stage, and a clear marker in the ERP of when it has its effect. For this reason we chose as one of the tasks in our experiments a task requiring lexical access: semantic categorization (living/non-living) in

Experiment 1, lexical decision (word/non-word) in Experiment 2. We assume (along with most psycholinguists) that the successive processes underlying semantic categorization are:

pre-lexical encoding of the stimulus in the form of an orthographic representation of the identity and order/position of its constituent letters;

lexical access, which comprises *lexical identification* —matching of the orthographic representation to stored word forms, and *semantic activation* — activation of the meaning or semantic features associated with the best-matching form(s);

response selection (or decision) — comparison of the activated meanings or semantic features to some template of the response categories or their boundary and selection of one of the two response categories;

response execution — translation of the abstract response category into an overt motor response.

Note that we do not assume these processes or sub-processes to be discrete: they may overlap, and lexical identification and semantic access may (or may not) interact.

The processes required for the lexical decision task are similar. Although finding a good match in the lexicon of word-forms is logically sufficient for a word decision, semantic activation provides additional evidence (because only words have meanings). Response selection for lexical decision presumably involves discriminating the states of activation of lexical/semantic representations triggered by words and non-words.

Frequency of use of a word in the language has a considerable — and similar — impact on semantic categorization time and on lexical decision time (Monsell, Doyle, & Haggard, 1989). Although there has been some argument about the locus of the frequency effect (Balota & Chumbley, 1984; Monsell et al., 1989), all models of lexical access assume more rapid matching or activation for more familiar word forms and more rapid activation of their meanings

(e.g., Norris, 2006). The effects of frequency can be seen early in the ERP waveform. Sereno, Rayner and Posner (1998) reported a larger N1 with a concurrent left anterior positivity for low frequency (LF) than for high frequency (HF) words 132-192 ms after the stimulus onset. More negative amplitudes for LF than for HF words were observed by Proverbio, Vecchi and Zani (2004) at 150–190 ms, and at 140–200 ms by Dambacher, Kliegl, Hofmann and Jacobs (2006). King and Kutas (1998) observed what they refer to as "frequency sensitive negativity" (FSN), which has a left frontal topography and peaks at ~280 ms for HF and ~335 ms for LF words. Later in the ERP it is well established that N400 amplitudes are larger for low compared to high frequency words (see Barber & Kutas, 2007, for a review). The earliest of these effects is controversial, but it appears that from 200 to 300 ms after visual presentation of a word, ERP markers of effects of word frequency can be reliably detected.

The only previous attempt to partition the RT cost of a task-switch using ERPs of which we are aware is Hsieh and colleagues' (Hsieh, 2006; Hsieh & Liu 2005; Hsieh & Yu, 2003) use of P3 and lateralized readiness potentials (LRPs). Their finding (Hsieh, 2006) was that P3 (which they regarded as a marker of stimulus identification) was not delayed by a task-switch, while the onset of the LRP (measured from stimulus onset) was delayed by a task-switch. (LRP onset measured relative to the response was unaffected). Given the standard view that LRPs, which begin to build up about 150 ms before the response is detected, correspond to motor activation of the corresponding hand – i.e. response execution, the latter finding appears merely to confirm the expectation of almost all theorists that, when the same response set is used for both tasks, task switching prolongs response selection *or earlier* processes. The failure to find a delayed P3 might be taken to suggest that processes earlier than response selection are not prolonged. However, P3 is sensitive to a number of factors, e.g., saliency or frequency of a

target, and its magnitude may be modulated by the relative saliency of a switch. Even if one found a delay in P3 one would not be able to uniquely specify which process is delayed.

Our approach is somewhat different. In Experiment 1, we used the onset of the effect of word frequency in the ERP as a marker to detect a difference in processing specific to a processing stage in the task of interest and orthogonal to switching. A delay in the onset of the frequency effect on task-switch trials would suggest that the switch prolongs pre-lexical encoding or an early phase of lexical access. In Experiment 2 we used the onset of the effect of lexicality — whether the item was a word or not — in the same way.

Choice of the other task. The second task was chosen to minimize the likelihood that significant reorientation or redistribution of spatial attention would be required when switching between that task and the lexical task. For example, attention to a single letter in a word produces a gradient of spatial attention different from the effect of attention to the whole word (LaBerge, 1983), as well as modulating the Stroop effect (Besner & Stolz, 1999). Processing in extrastriate cortex indexed by P1/N1 is substantially modulated by shifts in spatial attention (e.g., Hopf & Mangun, 2000). Hence we needed a second task that would require a distribution of attention over the whole letter string similar to that induced by word recognition (LaBerge, 1983). To this end, the stimuli were short words (in both experiments) and non-words (in Experiment 2) printed in a mixture of red and blue letters. For the second task, participants had to decide whether these colors were distributed symmetrically (e.g. RBRBR) or asymmetrically (e.g. RRBBR) about the center. To perform the lexical task in each experiment required attention to orthographic properties of the letter string (letter identity and order), and the coloring of the letters was irrelevant. To perform the symmetry task required attention to the spatial pattern of colors across the string, and the lexical content was irrelevant.

Lexical activation during performance of the symmetry task

For the symmetry task, processing of lexical content was not merely irrelevant; it could be a nuisance due to activation of the inappropriate response. Evidence of involuntary word recognition (e.g., from the Stroop task) nevertheless led us to expect to see ERP markers of lexical access on the symmetry task trials, and these are of interest for two reasons.

First, involuntariness of lexical access in a skilled reader remains a controversial issue. It is well known that the Stroop effect is substantially attenuated if the color name and the word are spatially separated (Kahneman & Henik, 1981), or if the participant attends to a single letter rather than the whole letter string (Besner, Stolz, & Boutilier, 1997). However, with a reading-appropriate distribution of spatial attention over the letter string (LaBerge, 1983) involuntary access may still be the default assumption. Taken seriously, this would imply that provided the locus and distribution of spatial attention is appropriate, lexical access will occur unmodified by whether the task on the current trial requires it or not. On the other hand, as we note above, it is uncontentious that people can attend to dimensions as well as locations and objects (Treisman, 1969; Müller, Reimann, & Krummenacher, 2003); hence the alternative possibility is that to perform the non-lexical task participants can modulate, without changing the spatial "spotlight", the degree to which orthographic properties are attended and lexical representations activated, not just later processes. The presence and amplitude of frequency- or lexicality-related effects in the ERP on the symmetry task trials should provide relevant evidence. We note that experiments combining lexical tasks with PRP (Besner, Reynolds, & O'Malley, 2009; Ruthruff, Allen, Lien, & Grabbe, 2008), attentional blink (Vachon & Jolicoeur, 2011), or subliminal priming (e.g., Kiefer, 2002) manipulations have led to a wide range of opinion on whether lexical access requires attention (in the sense of the allocation of central resources rather than spatial attention).

If it does, attending to the colors of the letters rather than the word they spell would be expected to attenuate or eliminate markers of lexical access.

A second reason for examining the effects of lexical variables on the symmetry task ERPs is because the task-set inertia account of the residual cost predicts that, if it is the stimulus-set (rather than, or instead of the response-set) that carries over, from one trial to the next to interfere with the required task-set, we should see more evidence of lexical access on the switch than on repeat trials. So far the only neurophysiological evidence of this kind comes from Yeung et al.'s (2006) fMRI study, as reviewed above: greater activation of a "face area" was associated with greater difficulty of switching to the word task, and of a "word area" with greater difficulty of switching to the face task. However, with a CSI of only 450 ms, it is not clear that their RT cost was fully "residual", and the fMRI signal was generated by sequences of four trials in each task, of which only the first could be a switch trial. The temporal precision of ERP allows us to look more precisely for trial-by-trial modulation of processing appropriate to the currently irrelevant task.

Experiment 1

The stimulus was a word whose letters were colored red or blue. The participant was instructed, by a visual cue, either to classify the word as referring to a living or nonliving object, or to classify the distribution of colors over the letters as symmetrical or asymmetrical. The main questions addressed were: (1) is the evolution of the frequency effect in the ERP on the semantic categorization trials delayed on a switch trial in such a way as to suggest that a task-switch prolongs early (lexical or pre-lexical) or later (response selection) processes? (2) Are markers of lexical access detectable on symmetry task trials, and does a task switch amplify them as a task-set inertia account would suggest?

The symmetry task was designed as a Go-Nogo task in the hope of using the N2 component in the ERP as a marker of completion of the symmetry judgment. Subjects responded to asymmetrically colored words with a key press and withheld the response for symmetrically colored ones. A substantial literature shows that in go/no-go discrimination tasks N2 and the longer-latency anterior-central P3 are both amplified on a no-go trial. Both effects have been proposed to reflect inhibition of the prepotent response (Eimer, 1993; Lavric, Pizzagalli, & Forstmeier, 2004); alternatively, N2 might reflect conflict between that activation of the response and withholding it (Nieuwenhuis, Yeung, & Cohen, 2004). Either way, since the symmetry judgment must largely precede the decision to respond or withhold the response, we hoped the N2 would help to localize the detection process within the RT. As it turned out, switching tasks eliminated the N2, so it could not be used as a marker in this way.

Although the focus of the present experiment was on post-stimulus ERPs we also examined ERPs during the cue-stimulus interval for evidence of active preparation for a task switch. Several switch-specific modulations of the pre-stimulus waveform have been found in preparation intervals, including a protracted posterior positive deflection starting from ~400 ms or more after the cue/previous response (e.g., Karayanidis, Coltheart, Michie, & Murphy, 2003; Lavric, Mizon, & Monsell, 2008; Rushworth, Passingham & Nobre, 2002), an anterior negativity in the same latency range (e.g., Astle, Jackson, & Swainson, 2008; Lavric et al., 2008), a longer-latency anterior-central negativity (Tieges, Snel, Kok, Wijnen, Lorist, & Ridderinkhof, 2006; Tieges, Snel, Kok, Plat & Ridderinkhof, 2007), and an early anterior positivity (Lavric et al., 2008; Rushworth et al., 2002). The first two components have been linked to advance task-set reconfiguration, because their amplitude co-varied with the effectiveness of preparation over trials (within subjects, Karayanidis, Provost, Brown, Paton, & Heathcote, 2011; Lavric et al., 2008), as well as over subjects (Elchlepp et al., 2012; Lavric et al., 2008).

Method

Participants. Eighteen right-handed students from the University of Exeter (15 female, 3 male) aged between 18 and 38 (M=21) received two course credits for participation plus a monetary performance bonus (maximum £4 — see below).

Stimuli and Procedure. On each trial, a 3-7 letter word (subtending 1˚- 2.5˚ horizontally) was displayed centrally in a non-proportionally spaced font (Lucinda Console), with some letters colored blue and the rest red. The participant was asked to perform one of two tasks. The "semantic" task was to categorize the word as referring to a living or non-living object by pressing the right or left arrow key; the two categories were equiprobable. The "symmetry" task was to determine whether the two colors were distributed asymmetrically or symmetrically relative to the middle of the word. For asymmetrically colored words participants made a "Go" response with a key press (left arrow key for half of the participants and right arrow key for the other half). Symmetrically colored words required withholding any response ('Nogo' trials). In line with the conventional Go-Nogo ERP paradigm, the probability of Nogo trials (37.5%) was lower than that of Go trials (62.5%).

The sequence of tasks was unpredictable, with task identity indicated on every trial by one of four cues (two per task). A cue appeared 800 ms before the stimulus and remained on the screen with the stimulus until a response was made (see Figure 1A). Each cue was a pair of small shapes displayed above and below the word's center to provide a midline marker and fixation as well as specify the task; their close proximity to the word was intended to minimize eye movements (see Figure 1B). Each possible assignment of pairs of cues to tasks was used for 6 participants. To avoid confounding cue and task changes, the cue changed on every trial (which is why we needed two cues per task). The task changed on a random third of trials; this relatively low probability was intended to discourage subjects from preparing for a possible

switch prior to cue presentation (cf. Monsell & Mizon, 2006). On all the trials in the semantic task and on Go trials in the symmetry task, the stimulus was displayed until a response was made, timing out at 2500 ms; on Nogo trials the stimulus was presented for 1500 ms. If a key press was detected on Nogo trials, or the wrong key or no key was pressed on other trials, the word "ERROR" was displayed for 1500 ms; otherwise the interval between each response and the cue for the next trial was 800 ms.

<<INSERT FIGURE 1 ABOUT HERE>>

Of the 384 word stimuli used, half referred to living and half to non-living objects. Half of each subset consisted of high frequency nouns (HF: >10/million, mean log frequency 1.3), and half low frequency nouns (LF: <5/million, mean log frequency 0.3), selected using the CELEX database of frequencies in British English (Baayen, Piepenbrock, & Van Rijn, 1993). In each of those four subsets 37.5 % of the words had a symmetrical pattern of colors. Each word was presented three times throughout the experiment, always in the same task, with the cue assigned randomly. The assignment of word stimuli to tasks was counterbalanced over subjects, as was the combination of occurrence of the word (1st, 2nd or 3rd) with task transition (switch versus repeat). That is, a third of the subjects saw a subset of words (which contained equal numbers of living/non-living by LF/HF words) first on switch trials (and for the second and third time on repeat trials), another third saw those words on switch trials on the second occurrence, and the remaining third of subjects saw them on switch trials at third occurrence during the experiment. Hence, words could be in one of 12 sets (living/non-living x LF/HF x switch at 1st/2nd/3rd occurrence) and these sets were matched for word length distribution. Furthermore, mean log word frequency for symmetrically and asymmetrically colored words was matched within high and low frequency sets. Subject to these constraints, the order of trials was randomized anew for each participant.

The EEG session was preceded by practice comprising four single-task blocks of 24 trials each (two blocks per task), one block of 24 trials to teach cue-task associations, and two 49-trial blocks like the experimental blocks. In the practice session, Go trials timed out after 5000 ms. Following practice, participants performed 24 blocks of 49 trials during which EEG data were acquired, comprising in total 96 switch and 192 repeat trials for each task by frequency category plus a warm-up (first) trial in each block.

Participants were instructed to use the cue to prepare for the upcoming task. To encourage effective preparation a score (mean RT/10 + errors*5) was computed for each block and a bonus payment was made for blocks on which the score was lower than the running average of the previous blocks.

EEG acquisition and pre-processing. The EEG was sampled continuously at 500 Hz with a bandpass of 0.016-100 Hz, the reference at Cz and the ground at AFz using 64 Ag/AgCl active electrodes (ActiCap, Brain Products, Munich, Germany) connected to BrainAmp amplifiers (Brain Products, Munich, Germany). There were 61 electrodes on the scalp in an extended 10-20 configuration, one on each earlobe and one below the right eye. Electrode locations were adjusted using a CMS ultrasound digitizer (Zebris Medical, Isny, Germany) and their impedances kept below 10 kΩ. Following offline filtering with a 20 Hz low-pass filter (24 dB/oct) and re-referencing to the linked ears, the EEG was segmented into a 1600 ms epoch, time-locked to the cue and baseline-corrected relative to the average amplitude of the 100 ms preceding the cue. Segments associated with incorrect responses were discarded, as were segments following errors and those from the warm-up trials. The resulting segments were visually inspected and segments containing ocular, muscle, movement and other artifact removed. The remaining EEG segments were averaged for every participant and condition.

ERP latency analysis. The primary aim of our experiments was to determine whether and by how much certain landmarks in the ERP were delayed on switch trials, thus providing a basis for partitioning of the RT switch cost among different stages of processing. To identify such landmarks, we computed (separately for switches and repeats) the ERP difference wave corresponding to the manipulation associated with a process (or set of processes). In Experiment 1, the frequency difference wave was computed by subtracting the amplitude of the ERP for high-frequency words from the ERP for low-frequency words. We then compared the latency of peaks (minima and maxima) in the difference waves for switch and repeat trials. To circumvent the issue of selecting individual electrodes for this comparison, latency calculations were performed on the Global Field Power (GFP) of the difference wave, which is the square root of the mean of all scalp channels' squared amplitudes (see Hamburger & van de Burgt, 1991, on the superiority of GFP for latency analyses).

To assess any such latency delays statistically, we needed to determine not only the mean but also the variance over subjects of the position in time of a given landmark. Since individual subjects' waveforms or difference waves are typically too noisy to unambiguously locate every landmark seen in the grand average, we used the "jackknifing" method (Miller, Patterson, & Ulrich, 1998). In accordance with this procedure, given 18 subjects, as well as the grand average, a further 18 average waveforms and difference waves were computed, each omitting one subject's data. A peak or inflexion was estimated in the same way for each of these averages, and the standard deviation of its shift in temporal location, between switch and repeat trials, computed over the 18 sub-averages. From this, an estimate of the variability of the shift in location of the landmark in time over individual subjects could be derived using Miller et al.'s formula for computing the appropriate within-subjects standard error, and hence the t-statistic.

ERP amplitude analysis. To investigate the early effects of switch and lexical frequency on ERP amplitude we analyzed the 100-500 ms interval following stimulus onset. The 0-100 ms range was not included because there are no documented effects of a task switch this early except the 'spillover' of the switch-induced posterior positivity from the pre-stimulus interval, which does not index stimulus-related activity (Elchlepp, et al, 2012; Lavric et al, 2008). To ensure temporal specificity, amplitudes were averaged in four equal time-windows of 100 ms and submitted to separate ANOVAs. In addition, we also analyzed the late part of the CSI in the larger time-window of 500-800 ms following cue-onset, where previous research has documented robust and protracted effects of a task switch linked to advance TSR (see Introduction). To assess potential interactions of the experimental conditions with scalp topography whilst minimizing the number of levels of topographic factors, amplitudes were averaged for groups of electrodes along the anterior-posterior axis (4 levels) and laterality (3 levels) dimensions (see Figure S1 in the Supplementary Materials); the ANOVA included these *region* and *laterality* factors, as we shall call them. Significance levels were adjusted using the Huynh-Feldt correction for violations of sphericity (but unadjusted degrees of freedom are reported).

Results and Discussion

Behavioral results. Mean correct RTs and error rates for all combinations of switch/repeat, task, frequency and symmetry are shown in Figure 2 (see also Table 1). (The first trial of each block and trials preceded by errors were excluded from the behavioral and ERP analyses.) For the symmetry task, RT analyses could be run only for Go trials. A switch (2) by task (2) by frequency (2) ANOVA on RTs found significant main effects of switch, $F(1, 17) = 33.72$, $p < 0.001$, task, $F(1, 17) = 22.87$, $p < 0.001$ and frequency, $F(1, 17) = 7.01$, $p < 0.05$. Responses were slower on switch trials than on repeat trials, for low frequency than for high frequency

words, and in the semantic task than in the symmetry task. The mean and standard error of the switch cost were 50 ± 12 ms for the semantic task, and 52 ± 13 ms. for the symmetry task. The effect of frequency (LF versus HF) was greater in the semantic task (30 ms) than in the symmetry task (7 ms), though the difference was only marginally reliable, $F(1, 17) = 3.23$, $p = 0.09$.

A separate switch x symmetry x frequency ANOVA was run for the semantic task. Apart from the expected main effects of switch, $F(1, 17) = 16.48$, $p < 0.001$, and frequency, $F(1, 17) = 17.37$, $p < 0.001$, a reliable effect of symmetry, $F(1, 17) = 4.45$, $p = 0.05$ was found. Symmetrically colored words (which required withholding a response in the other task) were responded to more slowly (852 ms) than asymmetrically colored words (833 ms), suggesting some interference from the irrelevant task.

<<INSERT FIGURE 2 ABOUT HERE>>

An ANOVA on the error rates including the factors switch, task, frequency and symmetry found reliable effects of switch, $F(1, 17) = 21.3$, $p < 0.001$, frequency, $F(1, 17) = 7.17$, $p < 0.05$, and symmetry, $F(1, 17) = 18.75$, $p < 0.01$, a reliable interaction between task and symmetry, $F(1, 17) = 21.8$, $p < 0.001$, and a marginally reliable three-way interaction between switch, task, and frequency, $F(1, 17) = 3.29$, $p = 0.087$. To investigate these effects in more detail, separate ANOVAs were run for the two tasks. For the semantic task, the switch cost (1.5 $\pm 0.4\%$), $F(1, 17) = 10.66$, $p < 0.01$ and frequency effect (1.6 $\pm 0.6\%$), $F(1, 17) = 9.91$, $p < 0.01$ were significant. For the symmetry task, the switch cost (2.6 $\pm 0.8\%$), $F(1, 17) = 10.11$, $p < 0.01$ and symmetry (Go vs. Nogo) effect, 5.5 $\pm 1.1\%$), $F(1, 17) = 25.16$, $p < 0.001$ were significant, the latter reflecting more commission errors on Nogo trials (8%) than response omissions on Go trials (2.5%). Switching interacted with frequency, $F(1, 17) = 4.72$, $p < 0.05$, such that the error switch cost was larger for HF words (3.4%) compared to LF words (1.6%), suggesting more

automatic lexical access for HF words and hence greater interference with the processing of the relevant (symmetry) dimension.

This experiment was aimed at examining the 'residual' switch cost, i.e. the cost that is left after effective preparation. However, whether preparation was indeed effective can only be determined by comparing performance with a relatively long CSI (as in the present experiment) to that with a short CSI. Although it has been claimed that a reduction in switch cost is not achieved when CSI is manipulated across participants (e.g., Altmann, 2004), in our lab we have obtained strong preparation effects in between-participants manipulations of CSI. We therefore ran a short CSI version of the present experiment: 18 participants performed exactly the same paradigm, with ERP recording, the only difference being that the CSI was 200 instead of 800 ms and the response-cue interval was 1400 ms (to keep the response-stimulus interval the same). In the short-CSI (200 ms) experiment the overall switch cost was 113 ms, much greater than the 51 ms observed in the present long-CSI (800 ms) experiment. A between-experiment contrast showed that the reduction in switch cost with increasing CSI was highly significant $F(1, 34) = 15.16$, $p < 0.001$. The error switch cost also reduced from 3% to 2%, though not significantly so, $F(1, 34) = 1.84$, $p = 0.18$.

To sum up, behavioral results confirm that our manipulations had the desired effects: reliable effects of switching, frequency and Go-Nogo were observed. In addition, the comparison with the short CSI study indicates that participants were indeed using the 800 ms to prepare effectively so that the major part of the switch cost is "residual"[2].

Semantic task ERPs: latency analysis. The primary aim of the experiment was to use ERP to ask which processes are prolonged on a task-switch trial when the participant has had ample time to prepare for the switch. We selected the semantic categorization task so that the onset of the

---

[2] Of course, to be confident that the cost is fully "residual" one would need to determine its asymptote, which requires at least one extra longer CSI.

word frequency effect around 200-250 ms in the ERP could serve as a marker of lexical access.

Hence we could ask: does switching tasks prolong processes before, simultaneous with, or

following lexical access?

To answer this question we looked at the effect of a task-switch on the latency of the

frequency effect in the ERP, relative to the effect on the overall response time (see LF and HF

grand average waveforms and difference waves for switch and repeat in electrode Pz in Figure 3;

topographies of this difference are available in Figure 4B). RT in the semantic task averaged

815 ms on task-repeat trials, and was prolonged by 50 ± 12 ms for the first response following a

change of task. The frequency effect in the ERP began around 200 ms after the stimulus onset.

The morphology of its evolution appears delayed from the outset on switch trials: an initial

maximum in the GFP difference wave (see Fig. 3) around 275 ms was delayed by a not yet

reliable 26 ± 17 ms, $t(17) = 1.53$, p = 0.14. A following minimum at round 300 ms was delayed

by a reliable 42 ms ± 13, $t(17) = 3.17$, p < 0.01, and the peak of the frequency effect at 500 ms

by 62 ± 16 ms, $t(17) = 3.94$, p < 0.001[3]. These estimates of delay are sizeable relative to the 50

ms effect of a task switch on response latency. The delay begins to emerge at about 230 ms post-

stimulus and is sizeable and reliable around 300 ms. In relation to a mean RT of 815 ms, it

indicates that a substantial fraction of the overall effect of a task switch on RT occurs in or prior

to lexical access, and hence well prior to response selection, contrary to a number of

interpretations of the residual cost reviewed in the Introduction.

We had also hoped to use the latency of the N2 potential typically associated with Go-

Nogo trials to localize the switch cost with respect to processing in the symmetry task, but found

little sign of the N2 effect in the present data. (We have found a typical N2 Go-Nogo difference

---

[3] We observed reliable latency delays also in individual channels (e.g., Pz) but report the GFP analysis because it avoids the arbitrariness of picking particular channels and improves the signal-to-noise ratio in estimating peak latency (see Hamburger & van de Burgt, 1991).

in this task, and with these stimuli, when no switching was required. This finding — that task switching largely eliminated the N2 effect — is of interest but not germane to our present purpose. We report it elsewhere: Elchlepp, Lavric & Monsell, in preparation.)

<<INSERT FIGURE 3 ABOUT HERE>>

Pre-stimulus ERPs. Figure 4A shows pre-stimulus ERPs for both tasks in electrode Pz as well as the topographies of the switch-repeat difference across the whole cue to stimulus interval; the inferential statistics for the associated analyses (ANOVA with the factors switch, task, region and laterality) can be found in Table A1. We mention these here for two reasons. First, they showed a late positive-polarity deflection in the ERP ("posterior positivity") on switch trials (relative to repeats) – the component consistently linked with successful preparation (see Karayanidis, Jamadar, Ruge, Phillips, Heathcote, & Forstmann, 2010, for a review). For both tasks we found a robust and extended switch positivity from about 500 ms onwards, reflected in the significant main effect of switch, which was greatest in the posterior regions of the left hemiscalp (reflected by the significant interactions between switch, region and laterality). This provides additional reassurance that participants did use the CSI to prepare for a switch. Secondly, this extra positivity on switch trials towards the end of the cue-stimulus interval explains why there is an initial switch-repeat difference in the waveforms subsequent to the stimulus.

Post-stimulus amplitudes, semantic task. Figure 4B shows the ERPs in electrode Pz, along with the switch-repeat and LF-HF amplitude differences presented as maps over all scalp electrodes in 75-ms intervals (see more waveforms in selected electrodes in Figure S2 in the

<<INSERT FIGURE 4 ABOUT HERE>>

Supplementary materials); the associated inferential statistics (from a switch by frequency by region by laterality ANOVA) are presented in Table A2. As expected on the basis of previous

research (e.g., Dambacher et al., 2006, Proverbio et al., 2004) and as suggested by the difference waves in the latency analysis above, there was a substantial effect of lexical frequency on ERP amplitude – a widespread negativity for low-frequency items maximal over parietal electrodes – which emerged at ~200 ms following stimulus onset and was largest during and after the N400 component of the ERP. This effect was statistically significant in the 200-500 ms range (main effect of frequency). The magnitude of the frequency effect did not seem to be reliably modulated by switching, judging by the absence of significant interactions involving factors frequency and switch. The analysis also revealed effects of switch – first the spill-over of the pre-stimulus positivity (see above), then a switch-negativity with a central-posterior scalp distribution (cf. Elchlepp et al., 2012; Karayanidis et al., 2003) – both effects (the positivity in the 100-200 ms time-window and the negativity in the subsequent time-windows) were statistically significant, as reflected in the interactions between switch, region and/or laterality in all time-windows and the main effect of switch at 400-500 ms.

Post-stimulus amplitudes, symmetry task. Another aim of the experiment was to use the word frequency effect on symmetry task trials as an index of the "automatic" occurrence of lexical access in the symmetry task, when recognition of the word would be counterproductive. Figure 4C shows switch-related and frequency-related differences in the same way as for the semantic task (for a larger selection of electrodes from the symmetry task, see Figure S3 in the Supplementary materials); the associated inferential statistics are summarized in Table A3 (as for the lexical task, the factors were switch, frequency, region and laterality). The earliest effect of lexical frequency (in the 200-300 ms time-window) was manifest in more positive-going amplitudes for low vs. high frequency words in the fronto-central scalp regions – an effect not observed in the semantic task – and more negative in occipital regions of the scalp (reflected in a significant frequency by region interaction). The more typical effect of lexical frequency

observed in the semantic task (widespread posterior negativity) eventually emerged and was significant in the last time-window (400-500 ms). Neither effect significantly interacted with switch, though both appear larger on switch trials. As in the lexical task, effects of switch were manifest (and statistically detectable in interactions with region or laterality) both as a carry-over of the pre-stimulus switch positivity (at ~100-200 ms) and as a switch negativity in the subsequent time-windows (reliably so at 200-300 ms and 400-500 ms).

Comparing the effect of lexical frequency in the semantic and symmetry tasks. The presence of an effect of frequency in the symmetry task, in which lexical processing is not required and is likely to be a hindrance, suggests that lexical access (or at least identification) is involuntary. However, to assess whether it is also *invariant*, we need to compare the latency and magnitude of the lexical frequency effects in the two tasks. First, we compared the latency of the onset of the frequency effect in the two tasks in much the same manner as we did for switch versus repeat trials in the semantic task above. Figure 5 shows the low-frequency versus high-frequency difference waves (as global field power) for the semantic and symmetry tasks. At ~268 ms there is an initial positive peak in the semantic task difference wave which appears displaced to ~360 ms in the symmetry task difference wave. In the jackknife analysis this delay of 92 ± 14ms was significant, $t(17) = 6.59$, $p< .001$. But do these peaks reflect equivalent processes? We computed the scalp distributions (topographies) of the frequency effect in a time-window surrounding the earlier peak (230-280 ms) for both tasks and a time-window surrounding the later peak (350-400 ms) for the symmetry task. As shown in Figure S4, the frequency effect in the semantic task in the 230-280 ms window was a negative-polarity deflection for low relative to high frequency words over the central frontal and central parietal scalp regions. A very similar distribution was apparent for the symmetry task in the 350-400 ms window. The scalp distribution of the (non-significant and small) deflections at 230-280 ms in

the symmetry task is quite different: a positive-polarity deflection over the central scalp. The scalp distributions thus suggest that the waveform difference in Figure 5 is appropriately characterized as a delay in the onset of the frequency effect in the symmetry task rather than, say, a mere attenuation of the amplitude of the difference wave in the symmetry task.

<<INSERT FIGURE 5 ABOUT HERE>>

Second, to directly contrast the amplitude of the word frequency effect in the two tasks we ran an ANOVA with the extra factor 'task' (in addition to factors switch, frequency, region and laterality). Frequency interacted significantly with task at 200-300 ms, $F(1,17) = 4.45$, $p = 0.05$, and 400-500 ms, $F(1,17) = 4.41$, $p = 0.051$. Between 200 ms and 300 ms, in the symmetry task there was little sign of the typical lexical frequency effect (more negative amplitudes for LF versus HF) present in the semantic task: it emerged only later, and appears attenuated. Thus, the frequency effect on ERP was both delayed (by nearly 100 ms) and considerably reduced (by over 50%) in the non-lexical task, indicating that lexical identification, though involuntary, is not invariant.

Experiment 2

Experiment 1 showed that switching delayed the onset of the frequency effect in the ERP for the lexical task from about 250 ms on, and that a frequency effect of similar morphology could be seen during performance of the non-lexical task. In the second experiment, we sought to extend these observations with a different and potentially earlier ERP marker of lexical access: the divergence between word and non-word waveforms in the lexical decision task. We also avoided the complexities of the Go-Nogo responses used in Experiment 1 by making the symmetry task a binary classification like the lexical task.

The point at which ERPs diverge for words and non-words in the lexical decision task is known to be as early as or earlier than, frequency effects. The earliest divergence reported, by

Sereno et al. (1998), occurred around 110 ms, with amplitudes becoming more positive for non-words than for words in parietal regions. Most studies have found the effect to emerge from about 200 ms onwards (Dehaene, 1995; Hauk, Patterson, Woolams, Watling, Pulvermüller, & Rogers, 2006(a); Hauk, Davis, Ford, Pulvermüller, & Marslen-Wilson, 2006(b); Hinojosa, Martin-Loeches, & Rubia, 2001; Martin-Loeches, Hinojosa, Gomez-Jarabo, & Rubia, 1999). Hauk et al. (2006a) report an early (~200ms) anterior positivity for non-words versus words, which subsequently turned into a posterior negativity for non-words. Martin-Loeches et al. (1999) describe posterior amplitudes becoming more positive for words than for non-words around 280 ms as the "recognition potential" for words.

Like word frequency, lexical status should have effects at every stage of lexical access: non-words activate multiple form candidates, but unlike words, no one candidate receives sufficient activation for recognition; non-words may to some extent activate meanings but obviously much less so than words. The leading edge of the divergence between the word and non-word waveforms is likely to reflect the earliest stages of lexical identification. However, if we randomly picked words and non-words, we would be in danger of confounding lexical status with orthographic properties and hence detecting a difference in pre-lexical encoding processes (in which letter identity and order/position become represented). Hence, we controlled for bigram frequency.

In fact, we not only matched the words and non-words on bigram frequency, but also manipulated bigram frequency orthogonally to lexical status. However, bigram frequency (BF) is strongly correlated with neighborhood size[4], another variable believed to have effects on early stages of lexical identification, when lexical forms partially matching the stimulus compete (Andrews, 1989, 1992, 1997). We did not attempt the difficult task of disentangling effects of

---

[4] By *neighborhood size* we mean the total number of words that differ from the target word by just one letter – a conventional measure of orthographic neighborhood used also in the ERP studies we review below.

bigram frequency and neighborhood size. Hence we shall use the term *orthographic familiarity* when interpreting the effects of this ambiguous variable. If we see an ERP effect of orthographic familiarity preceding lexical status effects, we can be reasonably confident that we are picking up either pre-lexical processing or early stages of lexical identification. And, if that effect too were delayed by a task switch, that would suggest a very early locus for the effect of a task switch. Relatively little is known about the ERP correlate(s) of bigram (or n-gram) frequency. Hauk et al. (2006a), who were particularly interested in very early stages of lexical identification, reported first effects of bigram frequency (apparently confounded with neighborhood size) in the ERP at a latency of ~110 ms with more positive amplitudes for low versus high BF items in posterior electrodes. Between this effect and the effect of lexicality at about ~210 ms, they further observed a significant interaction between these two variables (at about ~158 ms). Two ERP studies (Braun, Jacobs, Hahne, Ricker, Hofmann, & Hutzler, 2006, and Holcomb, Grainger, & O'Rourke, 2002) have reported orthographic neighborhood size effects in the form of more positive amplitudes for items with small neighborhoods from about 200 ms onwards- a protracted effect maximal over the midline scalp which increased monotonically in posterior channels. We will compare the effect of our orthographic familiarity manipulation to these effects.

Following the logic of Experiment 1, the main point of the second experiment was to examine whether the ERP effect of lexicality i.e. the point at which the waveforms for words and non-words start to diverge, is delayed on switch compared to repeat trials. To examine a potential delay in even earlier processes we could also compare the onset of orthographic familiarity effects on switch in relation to repeat trials. As in Experiment 1, it was important that the stimulus features and the allocation and spread of spatial attention did not differ substantially for the two tasks. Hence, the symmetry task continued to be a suitable second task. Also, as in

Experiment 1, the ERPs from this task provide clues about the extent to which orthographic processing and lexical access occur "automatically" on the symmetry task trials, and whether more so on task-switch trials, as an attentional interpretation of task-set inertia would suggest.

Method

Participants. Eighteen right-handed students from the University of Exeter (14 female, 4 male) aged between 18 and 45 (M=21) were given course credits for participation supplemented by a bonus payment (max £6) calculated on the basis of their performance (see below).

Stimuli and Procedure. On each trial, a 4-7 letter string was displayed, with some letters colored blue and the rest red, just as in Experiment 1. One task was lexical decision: categorize the letter string as a word or non-word by pressing the left or right arrow key, respectively; words and non-words were equiprobable. The other task was the symmetry task, now performed as a binary classification using the same keys: left for symmetrical, right for asymmetrical, the two being equiprobable.

The structure of stimulus sequences, the time-course of individual trials and the cues were the same as in Experiment 1. Of the 2400 items used, half were words and half non-words. Items were selected using the online version of the English Lexicon Project (ELP, Balota et al., 2007)[5]. Half of each subset consisted of high bigram frequency items (high-BF, ranging between 2200 – 6900, mean 3387; mean for words 3917; mean for non-words 2859) and half of low bigram frequency items (low-BF, ranging between 1 – 2000, mean 1270; mean for words 1493; mean for non-words 1047; Balota et al., 2007). BF by position "is based on the bigram frequencies that are sensitive to position within words. Position-sensitive bigram counts take into account the letter positions where a bigram occurs. For example, the bigram frequency for

---

[5] The ELP is an American English database used because it is – to our knowledge – the only database listing bigram frequency by position for a substantial number of words and non-words. We are aware that the BF measure used in this database is relatively crude because it does not take the length of the letter string into account.

DO in DOG counts DO bigrams only when they appear in the first two positions of the word."
(Balota et al., 2007, p.450). Both high-BF words and non-words had an average orthographic
neighborhood size (computed as the number of words that were different in one letter) of 3; low-
BF words and non-words had an average orthographic neighborhood of 1 (this difference in
orthographic neighborhood size was reliable, $t(2398) = 16.1$, $p < 0.001$). Both high- and low-BF
words had an average log word frequency of 2.3 (ranging from 0.3 to 5.3; Brysbaert & New,
2009). For further details of item characteristics see Table A4 in the Appendix. Each item was
presented only once in the whole experiment. Items were assigned to task and trial type
(switch/repeat) in a pseudo-random fashion for each participant, ensuring that over 6 participants
each item occurred once on a switch trial in each task and twice on a repeat trial in each task.
Half the items for each combination of BF, lexical status, switch/repeat and task had symmetrical
color patterns randomly assigned to them, half asymmetrical.

Participants were tested on two days no more than three days apart. On the first day, the
EEG session was preceded by practice comprising four single-task blocks of 24 trials each (two
blocks per task), one block of 24 trials for participants to learn the cue-task associations, and two
blocks of 48 trials of task-cuing trials like the experimental blocks. The electrode cap was then
applied and EEG data acquired during 20 blocks, each of 60 experimental trials preceded by a
warm-up trial. On the second testing day EEG data were acquired in further 20 blocks.

In both sessions subjects were instructed to use the cue to prepare for the upcoming task.
To encourage effective preparation the same incentive payment scheme was used as in
Experiment 1.

EEG and ERPs. The EEG set-up and analysis were the same as in Experiment 1, except that
the caps used in Experiment 2 lacked the CPz midline channel used in Experiment but contained
two extra lateral posterior channels (PO5 and PO6). The electrodes were regrouped for analysis

in the following regions: parietal left, now including CP3, CP5, TP7, P3, P5, parietal middle, now with CP1, CP2, P1, Pz, P2, parietal right with CP4, CP6, TP8, P4, P6; occipital left P7, PO5, PO7, O1; and occipital right P8, PO6, PO8, O2. The other regions remained as in Experiment 1.

On the first testing day electrode positions of a selection of electrodes (Fp2, F3, Cz, P5, Cp2, O2, T8) were recorded using a CMS ultrasound digitizer (Zebris Medical, Isny, Germany), so that they could be adjusted on the second day to ensure that electrode positions on the head remained constant across sessions.

Results and Discussion

Behavioral results. Mean correct RTs and error rates (excluding the first trial of each block and trials preceded by errors) for different conditions are plotted in Figure 6 (see also Table 2; effects of response congruence by task and switch-repeat are given in Table A5 in the Appendix). The ANOVAs on mean RTs and error rates included the factors session (first or second), trial type (switch, repeat), task (lexical, symmetry), lexicality (word, non-word), bigram frequency (low, high) and response congruency (congruent, incongruent).

Participants responded faster in the second (857 ms) than in the first (989 ms) session $F(1, 17) = 43.68$, $p < 0.001$. Responses were slower (955 ms) and error rates higher (7.9%) on switch trials than on repeat trials (891 ms; 5.7%) $F(1, 17) = 34.22$, $p < 0.001$, $F(1, 17) = 22.95$, $p < 0.001$. In the symmetry task responses were slower (990 ms) but more accurate (5.2%) than in the lexical decision task (856 ms; 8.4%) $F(1, 17) = 19.11$, $p < 0.001$, $F(1, 17) = 16.89$, $p < 0.01$. The RT switch cost was somewhat greater in the symmetry task (68 ± 14 ms) than in the lexical decision task (59 ± 12 ms), though this difference did not approach statistical significance, $p > 0.1$. The error switch cost on the other hand was larger in the lexical task ($3.3 \pm 0.7\%$) than in the symmetry task ($1.1 \pm 0.5\%$), $F(1, 17) = 9.71$, $p < 0.01$.

<<INSERT FIGURE 6 ABOUT HERE>>

Lexicality interacted with BF for RTs $F(1, 17) = 5.97$, $p < 0.05$, and errors $F(1, 17) = 6.04$, $p < 0.05$; responses to high-BF words were 8 ms faster and more accurate (6.4%) than to low-BF words (7.4%), while responses to high-BF non-words were 11 ms slower and more error prone (7.1%) than to low-BF non-words (6.1%). This lexicality by BF interaction was (marginally) reliable for lexical decision RTs $F(1, 17) = 4.3$, $p = 0.054$ and errors $F(1, 17) = 9.78$, $p < 0.01$, but not for the symmetry decision $F<1$ (RTs and errors); the interaction between BF, lexicality and task was not statistically significant for RTs ($F<1$), but it was for errors $F(1, 17) = 5.45$, $p < 0.05$.

Lexicality interacted reliably with task for RT $F(1, 17) = 69.74$, $p < 0.001$ and errors, $F(1, 17) = 33.33$, $p < 0.001$, with large word- non-word RT differences in the lexical decision task (101 ± 12 ms) and virtually no effect of lexicality in the symmetry task (1 ± 10 ms). More errors occurred for words (9.2%) than for non-words (7.5%) in the lexical task, whilst in the symmetry task slightly more errors were made for non-words (5.9%) than for words (5.4%). RTs for the two tasks also differed in effects of BF $F(1, 17) = 8.99$, $p < 0.01$, with responses 13 ms shorter to low (versus high) BF items in the lexical decision task (an effect entirely due to non-words) and 10 ms shorter to low (versus high) BF items in the symmetry task.

Responses were (marginally) faster and more accurate for response-congruent (917 ms, 4.1%) than -incongruent trials (929ms; 9.4%) $F(1, 17) = 4.37$, $p = 0.052$, $F(1, 17) = 42.72$, $p < 0.001$. For errors, congruency effects (incongruent minus congruent difference) were larger for switch (7.8%) than for repeat (2.8%) trials $F(1, 17) = 28.96$, $p < 0.001$ and larger in the lexical (7.2%) than the symmetry (3.4%) task $F(1, 17) = 12.58$, $p < 0.01$. The difference in congruency effects between switch and repeat was greater in the lexical task (switch 10.7%, repeat 3.7%)

than in the symmetry task (switch 4.9%, repeat 1.9%) (switch by task by congruency interaction, $F(1, 17) = 8.18$, p < 0.05).

To sum up, behavioral performance was as expected for both task switching and lexical decision with slower responses (and higher error rates) on switch compared to repeat trials and a substantial effect of lexicality on RTs and errors in the lexical decision task but not in the symmetry task. In the lexical decision task, lexicality interacted with orthographic familarity with longer RTs and more errors for low BF than for high BF words, and — the opposite pattern — shorter RTs and fewer errors for low BF than for high BF non-words. Response congruency effects indicated greater activation of responses via the irrelevant task-set on switch than on repeat trials.

Lexical decision task, ERP latency analyses. In the same way as for frequency in Experiment 1, grand average difference waves for the effect of lexicality were obtained, for switch and repeat trials separately, by subtracting the waveform for words from that of non-words and computing the global field power of the difference wave. As can be seen in Figure 7, no local maxima or minima (peaks or troughs) like those observed in the frequency difference waves in Experiment 1 were evident for use as landmarks in the latency analysis. Instead, the lexicality difference wave showed a monotonic increase with an onset that appears delayed on switch trials. To estimate the onset of the lexicality effect and compare its latency for switch and repeat trials, we fitted a bilinear function by least squares to an 80 ms (40 time samples, see Fig. 7) segment of each difference wave, using the inflection point between the best fitting two linear segments as an estimate of the onset. The second segment was chosen to be long enough to capture an initial portion over 40 ms (20 time samples) where the increase was approximately linear, but short enough that subsequent nonlinearities in the difference wave would not affect the estimate of the onset. For repeats, the whole segment fitted was 140 ms to 220 ms following stimulus onset.

For switches, the start and end point of the fitted segment was shifted by 40 ms (20 time samples, see Fig. 7) to capture the inflexion appropriately. As in Experiment 1, 'jackknifing' (Miller et al., 1998) was used to compute the standard error over subjects of the difference in onsets between the switch and repeat conditions.

<<INSERT FIGURE 7 ABOUT HERE>>

Figure 7 shows the GFP difference wave for the segments used for the bilinear function fitting and the GFP difference wave for the whole post-stimulus interval. When the lexical decision task was repeated, the (GFP) non-word-word difference wave started rising at ~182 ms. Switching to lexical decision delayed the onset of this lexicality effect by 59 ± 28ms, $t(17) = 2.11$, $p < .05$, a delay whose magnitude is similar to the RT switch cost in the lexical decision task (59 ± 12 ms). The earliness of this offset (~200 ms following stimulus onset) strongly supports the notion that a substantial part of the residual switch cost has an early locus.

Pre-stimulus amplitudes. Figure 8A shows topographies of the switch/repeat difference in pre-stimulus ERPs, and waveforms in electrode POz. An ANOVA on amplitudes averaged for the interval 500 – 800 ms following cue onset (see Table A6 for the relevant inferential statistics) confirmed previous findings: for both tasks there was a large late posterior positivity for the switch trials relative to the repeat trials maximal in the posterior electrodes and somewhat left-lateralised (significant main effect of switch and interactions between switch and region and/or laterality).

Post-stimulus amplitudes, lexical decision task. Scalp topographies of the effects of switching and lexicality, and waveforms in electrode Cz are presented in Figure 8B (a larger selection of electrodes can be found in Figure S5 in the Supplementary materials); the relevant inferential statistics are summarized in Table A7. The effects of switch on amplitude were relatively subtle (significant only in interaction with region and/or laterality) and likely confined to the spill-over

of the pre-stimulus positivity, given that they were largest in the earliest time-window. As one

can see in Figure 8B, the lexicality effect manifested itself first as frontal negativity for non-

words followed (from 400 ms onwards) by parietal negativity for non-words. These effects were

reflected in a significant main effect of lexicality from 200 ms onwards as well as significant

interactions of lexicality with region and/or laterality.

<<INSERT FIGURES 8 AND 9 ABOUT HERE>>

Orthographic familiarity affected amplitudes as early as 100 – 200 ms (see Fig. 9A and

Table A7) with more positive, left lateralized amplitudes for low versus high BF words (reflected

in a significant orthographic familiarity by laterality interaction). We had anticipated analyzing a

switch-induced delay in the onset of this orthographic familiarity effect in the same way as for

the onset of the lexicality effect. But although such an onset is clearly visible at about 150-200

ms on the task-repeat trials (see Fig. 9A), there is no clear onset on the task-switch trials. From

200 ms onwards orthographic familiarity interacted significantly with switch: low versus high

BF differences were large on repeat trials yet nearly absent on switch trials. Thus, switching

affected lexical decision ERPs in two ways: it delayed lexical processing as early as ~200 ms (as

shown by the latency analysis) and it almost abolished the large effects of orthographic

familiarity present on repeat trials from ~200 ms.

Post-stimulus amplitudes, symmetry task. ERP amplitude in the symmetry task was

influenced by lexicality throughout the whole interval (Fig. 8C and Table A8; more ERP plots

from the symmetry task are available in Figure S6 in the Supplementary materials). At 100-200

ms the lexicality effect appeared to be larger on switch than on repeat trials. However, it had

reversed polarity (more positive for non-words) and an earlier onset than in the lexical task,

which complicates interpretation. From 200 ms the lexicality effect (which now had the same

polarity as in the lexical task) was greater on repeat trials. Between 200 ms and 300 ms we also

observed a reliable effect of orthographic familiarity on switch but not on repeat trials, though the orthographic familiarity by switch interaction was not significant.

Comparing the effect of lexicality and orthographic familiarity in the two tasks. We started by comparing the onset latency of the lexicality effect in the two tasks. As for the switch repeat comparison, we estimated the onset of the effect by fitting a bilinear function. This resulted in an estimate of the onset of the lexicality effect (non-word difference) at 227 ms following stimulus onset in the lexical task. The effect was delayed in the symmetry task by 13 ± 12ms, but not reliably so, $t(17) = 1.15$, $p = .3$ (see Fig. 10).

To compare the amplitude of the effects of the lexical variables between tasks two additional ANOVAs were run including the factor task: (1) in addition to the factors switch, lexicality, region and laterality, and (2) in addition to switch, orthographic familiarity, region and laterality (there were insufficient trials for the full factorial analysis). The effect of lexicality was substantially larger (~3 times) in the lexical task than the symmetry task between 300 ms and 500 ms; 300-400 ms $F(1, 17) = 14.82$, p < .001, 400-500 ms $F(1, 17) = 22.3$, p < 0.001. The effect of orthographic familiarity differed reliably between tasks only in the 200 - 300 ms interval, $F(1, 17) = 4.94$, p < 0.05, with a larger difference in the lexical than the symmetry task. The lack of a more substantial difference in the later time windows is likely due to the feeble effect of orthographic familiarity on switch trials in the lexical task.

<<INSERT FIGURE 10 ABOUT HERE>>

General discussion

In two experiments we cued participants to perform a lexical classification or a non-lexical perceptual classification of a letter string, with 800 ms to prepare following the cue. As usual, RT was prolonged on a switch relative to a repeat trial. We used ERP latencies to investigate the locus of this switch cost within the latent interval for the lexical task. In Experiment 1,

participants semantically categorized words, and the onset and early evolution of the word frequency effect in the ERP provided a marker of lexical access. ERP waveforms for low and high frequency words started to diverge at ~220 ms. On switch trials the evolution of this frequency effect (reflected in its difference waveform) was delayed: a minimum at ~300 ms on repeat trials was postponed on switch trials by a significant 44 ms, a substantial proportion of the 50 ms RT switch cost. Experiment 2 confirmed and extended this finding using a different and earlier marker of the onset of lexical access: the effect of lexicality (word versus non-word). The onset of the lexicality effect at ~200 ms was delayed on switch trials by about 60 ms, again an amount similar to the switch cost. It is noteworthy that in Experiment 1, the frequency manipulation that generated the marker of interest in the ERP was orthogonal to the response category, whilst in Experiment 2 the lexicality manipulation aligned with the response category. The former design is perhaps preferable, as the response category is likely to be manifest in the ERP at some point, but 200 ms is implausibly early for response selection given the observed RTs.

In Experiment 2 we also manipulated (orthogonally to lexicality) bigram frequency; in our stimulus set this was correlated with neighborhood size, so we refer to this variable as orthographic familiarity. The effects of orthographic familiarity on ERP amplitude were modulated by a task switch in a qualitatively interesting way: on repeat trials BF effects were large and spread over wide areas of the scalp, on switch trials the effects were small and circumscribed.

Finally, the effects of lexical properties on ERPs during performance on the symmetry task indicated that unwanted but "automatic" lexical access occurred on the symmetry task trials. However, the magnitude of the ERP indices of lexical access was substantially reduced (by 2-3 times) when performing the symmetry task. There were some signs that this automatic access

was augmented immediately following a switch from the lexical to the symmetry task, e.g., the lexical frequency effects in the symmetry task were numerically larger on switch trials in Experiment 1, and the orthographic familiarity effects were more pronounced in the symmetry task on switch trials in Experiment 2, but one contrast was not consistent with this observation – the presence of a larger lexicality effect in the late part of the ERP on symmetry task repeat trials in Experiment 2. Hence, the present data neither confirm nor reject this particular prediction of inertia in an attentional or "pathway" component of task-set.

Implications for theories of the residual task-switch cost

The earliness of the effect of task switching on processing – between 200 and 300 ms post-stimulus — and the estimated sizes of the delay at this early stage — 42 ± 13 ms in Experiment 1, 59 ± 28 ms in Experiment 2, relative to the RT switch cost ($\sim$ 60ms in both experiments) — together suggest that a switch of task, after preparation, prolongs processing early within the latent interval – i.e. in early lexical or pre-lexical processing. We cannot conclude that there is *no* effect of a task switch on the later response-selection process but in these experiments the effect on RT appears to be much more due to the onset of response selection being delayed than to its duration being prolonged. How does this fit with the theories of the residual cost we reviewed in the Introduction?

Exogenously triggered completion of reconfiguration. Rogers and Monsell (1995) suggested a second stage of TSR that could not happen until triggered by the stimulus. Such a theory is compatible with an early locus of the residual switch cost, as the changed S-R rules must presumably be retrieved or activated in time for response selection.

Task-set inertia. There was clear evidence from frequency and lexicality effects in the ERPs in the symmetry task that lexical access occurred even when the participant was doing the symmetry task – a prerequisite for an inertia account (cf. Yeung et al., 2006). The early locus of

the effect of a prepared switch on the lexical task indicates that if the delay is due to task-set "inertia", it is *attentional* inertia — persisting allocation of attention (or processing resources) to the irrelevant stimulus attribute or processing pathway — that is primarily responsible, rather than persistent activation of the S-R mappings from the previous trial (which should specifically prolong response-selection on the switch trial). That is not to deny that there is persistence of the irrelevant S-R mapping on both switch and repeat trials — as manifest in the response congruence effect.

Associative reactivation of task-set. As conceived of by Waszak et al. (2003; 2004; 2005) the effect of extra associative reactivation of a task-set on a switch trial is to prolong response-selection. That does not seem compatible with an early locus of the effect of a switch indicated by the ERP latency analyses. But the early locus is compatible with the idea that what the stimulus reactivates, associatively, is a task-specific attentional orientation or resource allocation.

Response-contingent reconfiguration. Meiran's (2000) account of the residual cost was that while attention may be redirected to a different dimension during preparation for a task change, a response has to be generated to change the meaning of the response set (S-R weighting) and optimize performance; thus the residual cost reflects impaired response selection. Schuch and Koch (2003) also emphasize response-selection as the locus of the residual cost, proposing that even after full preparation a bivalent stimulus elicits conflict, and the residual switch cost reflects task-set suppression applied during response selection to resolve the conflict. If the task-set suppression (whatever it suppresses) is applied only during response selection, this will be too late to affect early processes. In contrast to both accounts, our results suggest it is largely processing prior to response-selection processing that is slowed on a task-switch trial. In their later CARIS model, Meiran et al. (2008) reversed their original conclusion to reach a compatible conclusion: attentional reweighting requires the presence of the stimulus (so it is not done in

advance) – hence generating a residual cost, while response remapping can be accomplished before stimulus onset.

In summary, our conclusion is that the ERP data from the lexical tasks demonstrate that, after preparation, the effect of a task-switch is mainly to prolong processes or insert extra processes well before response selection. For the lexical tasks that we used, the process prolonged could be pre-lexical encoding and/or relatively early stages of lexical access. Either is compatible with the idea of "attentional inertia": a difficulty reorienting attention to orthographic attributes, or a difficulty reactivating the lexical pathway, until the lexical task has been performed at least once.

Other evidence for attentional inertia

In most task-switching studies, the tasks differ in both the stimulus attribute to be attended to and the S-R rules to be applied, requiring a change in both "stimulus-set" and "response-set". But some studies labeled as "task-switching" have required switches between tasks that differ only in the attribute to be attended to. For example, Rushworth, Passingham and Nobre (2005) had participants respond with a constant S-R mapping to the character superimposed on one of two objects, the target object being defined by its color or shape. Switching (only) the dimension defining the target object incurred substantial switch costs, even with 2 seconds available for preparation. Meiran and Marciano (2002) report a switch cost for judging whether two multidimensional stimuli were the same or different, when what switched was which of the several stimulus dimensions had to match for a "same" response: the S-R rules mapping "same" and "different" to keys remained constant. Interestingly, this study showed no or only a weak reduction of switch cost with preparation for this effect of a cued dimension switch. In contrast, in an experiment reversing the response assignment on some trials, the switch cost did reduce with preparation: doubtless part of the motivation for Meiran et al's

(2008) revision of Meiran's (2000) account to conclude that the response set, but not the stimulus set (attention to a dimension), can be reconfigured prior to stimulus onset.

The visual search literature has extensively examined evidence both for passive priming of attention (revealed by costs in switching the feature or dimension defining the target from trial to trial), and top-down prioritization of the processing of a feature or dimension, revealed by benefits of cuing. For example, modest dimension-switching costs are seen in visual search for a single-feature singleton when the target changes between colour and shape singletons, and these costs are attenuated — but not eliminated — by verbally cuing the dimension of the target (Müller, Reimann, & Krummenacher, 2003). It is widely agreed that the detection of targets defined by specified features (Wolfe & Horowitz, 2004), dimensions (Krummenacher & Müller, 2012), objects, and even categories of object (Nako, Wu & Eimer, 2014) can be proritized through top-down control. Intense debate has focused on how early in processing this top-down control penetrates, with Theeuwes (2010), for example, contending that only attention to location can prioritize processing during the first 170 ms or so of feedforward sensory analysis, while others have argued, from the impact of task-set on RT and ERP measures of attentional capture by irrelevant singletons (Folk Remington, & Johnston's, 1992, "contingent capture") that the impact of top-down attentional set is earlier (e.g. Ansorge, Kiss, Worschech, & Eimer, 2011). What is less clear from this literature is the degree to which the impact of top-down control is limited in a way that we would describe as attentional inertia. It may be noted, however, that in the Müller et al (2003) study mentioned above, the cost of a dimension switch, although reduced by a cue, was not eliminated even with a generous 1550 ms CSI.

There are also long traditions of research on shifting attention between "channels" defined by location (Posner, 1980), ear (Treisman, 1971), modality (Spence & Driver, 1997), or object(s) in an array (Driver & Baylis, 1998; Duncan, 1984). Costs of switching attention

between channels and benefits of cuing the channel are well known, but the possibility of residual costs in intentional switching between channels does not seem to have been addressed as such until recent research by Koch and colleagues, who have applied cuing and CSI manipulations used in task switching to the intentional switching of auditory attention. Koch, Lawo, Fels, & Vorländer, (2011) presented a pair of digits dichotically, one in a male and one in a female voice (any of three voices per gender), preceded by a visual gender cue indicating which digit to classify as high or low. The participant could evidently use the cue to select the relevant digit, but a substantial switch cost was observed when the cued gender changed and (with two cues per gender to control for effects of cue priming) the cost showed no reduction at all when the CSI increased from 100 to 1000 ms. Lawo, Fels, Oberem and Koch (2014) compared gender-cuing to ear-cuing, and again found that the gender-cued RT and switch cost was insensitive to an increase in CSI, whereas the same opportunity for preparation substantially reduced ear-cued RT and reduced the switch cost (although the 3 way interaction was not reliable).

Combining contingent capture and task-switch paradigms, Lien, Ruthruff and Johnston (2010) have specifically argued that preparatory shifting of attention between feature targets leaves no residual cost. They cued participants to perform one of two classification tasks to one of four digits presented simultaneously in different colors; each classification was consistently associated with a specific color; the relevant color+classification was cued 1350 ms in advance of the stimulus by a single letter (e.g. 'R' for red and parity judgment). Lien et al. (2010) found that an irrelevant stimulus presented shortly before stimulus onset captured attention only when its color was task-relevant, and that this contingent capture effect was uninfluenced by task switch/repeat, suggesting that task-set preparation during the interval was completely effective in reconfiguring the feature set and thus suppressing capture by the irrelevant color.

Experiments in our lab (Longman, Lavric & Monsell, 2013; Longman, Lavric, Munteanu, & Monsell, 2014) have led to a different conclusion for spatial attention. Using tasks consistently associated with different spatial locations, and eye-tracking to index spatial attention, we found that task-switching induces both substantial delays in orienting appropriately, and a tendency to fixate the target location appropriate to the previous task, even with a generous interval for preparation. In another recent eyetracking investigation, Mayr, Kuhns and Rieter (2013) examined switching between discriminating the color and form of color and form singletons in a variant of visual search and found evidence for impaired attentional selection of a dimension on switch trials, even at long preparation intervals. These studies provide rather direct evidence for attentional inertia.

In summary, evidence of "attentional inertia" is beginning to accumulate from studies that have examined measures of intentional shifts of attention between locations/objects, features (perhaps), perceptual dimensions, and gender-defined auditory sources, though these are all "objects of attention" defined in a less complex way, or at a lower level of processing, than the lexical orthography versus color pattern attributes to which attention was required in the present experiments.

Possible limitations of the present study

Our experiments were intended to investigate the residual cost, i.e. the cost that is left after successful preparation. But it would be rash to claim that reconfiguration was completely accomplished before the stimulus on all trials. One of our previous investigations (Lavric et al., 2008) looked at preparation effects in trials with fast versus slow responses and found a switch positivity in the stimulus interval of slow-response trials similar to the one found on short CSI trials where there is no opportunity to reconfigure before the stimulus, suggesting that, even with

an interval sufficient for preparation, on a proportion of trials reconfiguration takes place only after stimulus onset (cf. De Jong, 2000, Nieuwenhuis & Monsell, 2002). Trial numbers in the present experiments did not allow us to perform separate analyses for fast and slow trials, so we cannot rule out the possibility that post-stimulus reconfiguration on a proportion of trials contributed to the processing delays measured on switch trials. We doubt, however, that delayed preparation is the sole or major source of the ERP latency effects observed, for the following reasons: the length of the CSI (800 ms) was chosen to be greater than the CSI at which switch cost becomes asymptotic in most studies; the residual cost in the lexical tasks (~60 ms) is of the same order as residual costs in other studies (e.g., Lavric et al., 2008; Nicholson, Karayanidis, Poboka, Heathcote, & Michie, 2005); the switch cost was reduced by about half in the 800 ms CSI experiment (Experiment 1) compared to the 200 ms CSI version, suggesting effective preparation); we found a large and highly reliable preparation-related switch-induced positivity in ERPs of both studies. These points suggest that participants used the 800 ms CSI on most of the trials to prepare, so that the switch cost we investigated was indeed largely the "residual" cost.

Another possible limitation is that, although we chose the tasks between which participants switched to require a similar distribution of spatial attention, we cannot be certain that the distribution was identical for the lexical and symmetry tasks. However, if substantially different distributions of attention were being achieved for the two tasks, we might expect participants to have been more successful in suppressing lexical access on symmetry trials (cf. Besner et al.'s modulation of Stroop interference by spatial zooming). We have recently obtained similar evidence for an early locus with a pair of tasks where reorientation of spatial attention is an even less likely source of the processing delay due to a task switch – vowel/consonant classification of a single letter and classification of its color; with other

precautions to avoid any shift in spatial orienting per se when the task changed (Elchlepp, Best, Lavric, & Monsell, in preparation).

Implications of the orthographic familiarity effects

We expected the effects of the orthographic familiarity variable to emerge early in the ERP, possibly even preceding the effect of lexicality, as bigram frequency, per se, is likely to influence pre-lexical processing of orthography. Local differences became apparent at ~100 ms but at this stage were not reliably modulated by switching. From about ~200 ms onwards they generalized across the whole scalp – but, in the lexical decision task, only for task-repeat trials (see Figure 9A and table A7 in the Appendix). This interaction with switching is unexpected and intriguing, because it reflects an overall reduction, rather than a delay, in the effect of orthographic familiarity during a task switch. Given the similarity of the effect of orthographic familiarity in timing and topography to previously reported effects of orthographic neighborhood size (e.g., Holcomb et al., 2002; Braun et al., 2006), it is likely that what we were seeing was in fact the effect of this variable.

But why does the effect appear only on task-repeat trials for the lexical decision task, but on the switch trials in the symmetry task (see Fig 9A versus 9B)? It is often claimed that to help them make a lexical decision, participants can monitor the global activation of the lexicon as a correlate of lexical status (e.g., Grainger & Jacobs, 1996). We could speculate that (a) this monitoring generates the ERP effect of orthographic familiarity we detect, (b) this monitoring takes a trial to re-establish after switching to the lexical decision task (so we see its ERP effect only on repeat trials for the lexical decision task), (c) that once established this monitoring tends to persists for a trial (so that we see the ERP effect on switch trials in the symmetry task), and (d) that the extra global activation of the lexicon generated, on average, by items with a large relative to low orthographic neighborhood does not, per se, generate a detectable signal in the

ERP, though attention to global activation does. This is, we acknowledge, quite a list of ad hoc assumptions.

Implications for the automaticity of word reading

As we reviewed in the introduction, claims about the automaticity of word recognition have, in recent years, yielded more nuanced but conflicting views, recognizing first the importance of the distribution of spatial attention (Besner et al., 1997) and secondly the results from dual-task, RSVP and priming studies of the resources required for lexical tasks. Although Besner and colleagues (Besner et al., 1997, 2009) have argued for the automaticity of at least early stages of lexical identification, others have demonstrated substantial modulation of ERP and priming indices of lexical or semantic activation as a function of a concurrent or preceding task (Kiefer & Martens, 2010; Lien, Ruthruff, Cornett, Godin, & Allen, 2008; Rabovsky, Alvarez, Hohlfeld, & Sommer, 2008; Vachon & Jolicoeur, 2011). Our own findings seem to be on a continuum with the latter group of studies. Unlike theirs, our participants had substantial time to recover from processing of a previous stimulus and prepare for lexical processing, but we still found that switching away from a different task postponed or prolonged early stages of lexical access. And ERP markers of lexical access occurred when performing the non-lexical task, albeit delayed and attenuated. This suggests that lexical access is not automatic in the sense of proceeding invariantly, even with appropriate allocation of spatial attention.

References

Ansorge, U., Kiss, M., Worschech, F., & Eimer, M. (2011). The initial stage of visual selection is controlled by top-down task set: New ERP evidence. *Attention, Perception, & Psychophysics*, 73, 113-122

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, 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.

Altmann, E. M. (2004). The preparation effect in task switching: Carryover of SOA. *Memory and Cognition*, 32, 153-163.

Altmann, E. M., & Gray, W. D. (2008). An integrated model of cognitive control in task switching. *Psychological Review*, 115, 602-639.

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 802-814.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, 234-254.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin and Review*, 4, 439-461.

Aron, A. R., Monsell, S., Sahakian, B. J., & Robbins, T. W. (2004). A componential analysis of task-switching deficits associated with lesions of left and right frontal cortex. *Brain*, 127, 1561-1573.

Astle, D. E., Jackson, G. M., & Swainson, R. (2008). Fractionating the cognitive control required to bring about a change in task: a dense-sensor event-related potential study. *Journal of Cognitive Neuroscience*, 20, 255-267.

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database. CD-ROM*. Philadelphia, PA.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchinson, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Brain Research Methods*, 39, 445-459.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 340-357.

Barber, H. A., & Kutas, M. (2007). Interplay between computational models and cognitive electrophysiology in visual word recognition. *Brain Research Reviews*, 53, 98-123.

Besner, D., Reynolds, M., & O'Malley, S. (2009). When underadditivity of factor effects in the Psychological Refractory Period paradigm implies a bottleneck: Evidence from psycholinguistics. *Quarterly Journal of Experimental Psychology*, 62, 2222-2234.

Besner, D., & Stolz, J. A. (1999). Context dependency in the Stroop paradigm: When are words treated as nonlinguistic objects? *Canadian Journal of Experimental Psychology*, 53, 374-380.

Besner, D., Stolz, J. A., & Boutilier, C. (1997). The Stroop effect and the myth of automaticity. *Psychonomic Bulletin and Review*, 4, 221-225.

Braun, M., Jacobs, A.M., Hahne, H., Ricker, B., Hofmann, M., & Hutzler, F. (2006). Model-generated lexical activity predicts graded ERP amplitudes in lexical decision. *Brain Research*, 1073-1074, 431-439.

Braver, T. S., Reynolds, J. R., & Donaldson, D. I. (2003). Neural mechanisms of transient and sustained cognitive control during task-switching. *Neuron*, 39, 713-726.

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.

Brown, J. W., Reynolds, J. R., & Braver, T. S. (2007). A computational neural model of fractionated conflict-control mechanisms in task-switching. *Cognitive Psychology*, 55, 37-85.

Bryck, R. & Mayr, U. (2008). Task selection cost asymmetry without task switching. *Psychonomic Bulletin and Review*, 15, 128-134.

Brysbaert, M. & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41, 977-990.

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*, 1084, 89-103.

Dehaene, S. (1995). Electrophysiological evidence for category-specific word processing in the normal human brain. *NeuroReport*, 6, 2153-2157.

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.

Driver, J., & Baylis, G. C. (1998). Movement and visual attention: The spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 448–456.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113, 501–517.

Eimer, M. (1993). ERPs elicited by Go and Nogo stimuli: Effects of attention and stimulus probability. *Biological Psychology*, 35, 123-138.

Elchlepp, H., Lavric, A., Mizon, G. A., & Monsell, S. (2012). A brain-potential study of task-switching with stimuli that afford only the relevant task. *Human Brain Mapping*, 33, 1137-1154.

Elchlepp, H., Best, M., Lavric, A., & Monsell, S. (in preparation). Partitioning the cost of a task switch with brain potentials reveals early attentional locus.

Elchlepp H., Lavric A., & Monsell, S. (in preparation). Switching abolishes the Nogo-N2.

Folk, C. L., Remington, R. W., & Johnston, C. J. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception & Performance*, 18, 1030-44.

Forrest, C.L., Monsell, S., & McLaren, I. P. L. (2014). Is performance in task-cuing experiments mediated by task set selection or associative compound retrieval? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 40, 1002-1024.

Gilbert, S. J., & Shallice, T. (2002). Task switching: A PDP model. *Cognitive Psychology*, 44, 297-337.

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. M. (1996). Orthographic processing in visual word recognition: a multiple read- out model. *Psychological Review*, 103, 518–565.

Hamburger, H. L. & van d. Burgt, A. G. (1991). Global Field Power measurement versus classical method in the determination of the latency of evoked potential components. *Brain Topography*, 3, 391-396.

Hauk, O., Patterson, K., Woollams, A., Watling, L., Pulvermüller, F. & Rogers, T. T. (2006a). [Q:] When would you prefer a SOSSAGE to a SAUSAGE? [A:] At about 100 ms. ERP correlates of orthographic typicality and lexicality in written word recognition. *Journal of Cognitive Neuroscience*, 18, 818-832.

Hauk, O., Davis, M.H., Ford, M., Pulvermüller, F. & Marslen-Wilson, W.D. (2006b). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *NeuroImage*, 30, 1383-1400.

Hinojosa, J. A., Martin-Loeches, M., & Rubia, F. J. (2001). Event-related potentials and semantics: An overview and an integrative proposal. *Brain and Language*, 78, 128-139.

Holcomb, P.J., Grainger, J., & O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience*, 14, 938-950.

Hsieh, S. (2006). The lateralized readiness potential and P300 of stimulus-set switching. *International Journal of Psychophysiology*, 60, 284-291.

Hsieh, S., & Liu, L.-C. (2005). The nature of switch cost: task set configuration or carry-over effect? *Cognitive Brain Research*, 22, 165-175.

Hsieh, S., & Yu, Y. T. (2003). Switching between simple response-sets: inferences from the lateralized readiness potential. *Cognitive Brain Research*, 17, 228-237.

Hopf, J. M., & Mangun, G. R. (2000). Shifting visual attention in space: An electrophysiological analysis using high spatial resolution mapping. *Clinical Neurophysiology*, 111, 1241-1257.

Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 181-211). Hillsdale, NJ: Erlbaum.

Karayanidis, F., Coltheart, M., Michie, P. T., & Murphy, K. (2003). Electrophysiological correlates of anticipatory and post-stimulus components of task switching. *Psychophysiology*, 40, 329-348.

Karayanidis, F., Jamadar, S., Ruge, H., Phillips, N., Heathcote, A, & Forstmann, B. U. (2010). Advance preparation in task-switching: converging evidence from behavioral, brain activation, and model-based approaches. *Frontiers in Psychology*, 25, 1-13.

Karayanidis, F., Provost, A., Brown, S., Paton, B., & Heathcote, A. (2011). Switch-specific and general preparation map onto different ERP components in a task-switching paradigm. *Psychophysiology*, 48, 559-568.

Kiefer, M. (2002). The N400 is modulated by unconsciously perceived masked words: Further evidence for a spreading activation account of N400 priming effects. *Cognitive Brain Research*, 13, 27-39.

Kiefer, M.,, & Martens, U. (2010). Attentional sensitization of unconscious cognition: Task sets modulate subsequent masked semantic priming. *Journal of Experimental Psychology: General*, 139, 464-489.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A., & Koch, I. (2010). Control and Interference in Task Switching—A Review. *Psychological Bulletin*, 136, 849-847.

King, J. W., & Kutas, M., (1998). Neural plasticity in the dynamics of human visual word recognition. *Neuroscience Letters*, 244, 61–64.

Koch, I., & Allport, A. (2006). Cue-based preparation and stimulus-based priming of tasks in task switching. *Memory & Cognition*, 34, 433-444.

Koch, I., Gade, M., Schuch, S., & Philipp, A. M. (2010). The role of inhibition in task switching – A review. *Psychonomic Bulletin and Review*, 17, 1-14.

Koch, I., Lawo, V., Fels, J., & Vorländer, M. (2011). Switching in the cocktail party – Exploring intentional control of auditory selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 1140-1147.

Krummenacher, J., & Müller, H. J. (2012). Dynamic weighting of feature dimensions in visual search: behavioral and psychophysiological evidence. *Frontiers in Psychology*, 3, 221.

LaBerge, D. (1983). Spatial extent of attention to letters and words. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 371-379.

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

Lavric, A., Mizon, G. A., & Monsell, S. (2008). Neurophysiological signature of effective anticipatory task-set control: a task-switching investigation. *European Journal of Cognitive Neuroscience*, 28, 1016-1029.

Lavric, A., Pizzagalli, D., & Forstmeier, S. (2004). When 'go' and 'nogo' are equally frequent: ERP components and cortical tomography. *European Journal of Neuroscience*, 20, 2483-2488.

Lawo, V., Fels, J., Oberem, J., & Koch, I. (2014). Intentional attention switching in dichotic listening: Exploring the efficiency of nonspatial and spatial selection. *Quarterly Journal of Experimental Psychology*. doi: 10.1080/17470218.2014.898079 (online first publication)

Lien, M. C., Ruthruff, E., & Johnston, J. C. (2010). Attention capture with rapidly changing

    attentional control settings. *Journal of Experimental Psychology: Human Perception and*

    Performance, 36, 1-16.

Lien, M.-C., Ruthruff, E., Remington, R. W., & Johnston, J. C. (2005). On the limits of advance

    preparation for a task switch: Do people prepare all the task some of the time or some of

    the task all the time? *Journal of Experimental Psychology: Human Perception and*

    Performance, 31, 299-315.

Lien, M., C., Ruthruff, E., Cornett, L., Goodin, Z., & Allen, P., A. (2008). On the

    nonautomaticity of visual word processing: Electrophysiological evidence that word

    processing requires central attention. *Journal of Experimental Psychology: Human*

    Perception and Performance, 34, 751-773.

Logan, G. D., & Bundesen, C. (2003). Clever homunculus: Is there an endogenous act of control

    in the explicit task-cuing procedure? *Journal of Experimental Psychology: Human*

    Perception and Performance, 29, 575-599.

Longman, C. S., Lavric, A., & Monsell, S. (2013). More attention to attention? An eye-tracking

    investigation of selection of perceptual attributes during a task switch. *Journal of*

    Experimental Psychology: Learning, Memory, and Cognition, 39, 1142-1151.

Longman, C. S., Lavric, A., Munteanu, C., & Monsell, S. (2014). Attentional inertia and delayed

    orienting of spatial attention in task-switching. *Journal of Experimental Psychology:*

    Human Perception and Performance, 40, 1580-1602.

Los, S. A. & van der Burg, E. (2010). The origin of switch costs: Task preparation or task

    application? *Quarterly Journal of Experimental Psychology*, 63, 1895-1915.

Martin-Loeches, M., Hinojosa, J. A., Gomez-Jarabo, G., & Rubia, F. J. (1999). The recognition

    potential: An ERP index of lexical access. *Brain and Language*, 70, 364-384.

Masson, M. E. J., Bub, D. N., Woodward, T. S., & Chan, J. C. K. (2003). Modulation of word-reading processes in task switching. *Journal of Experimental Psychology: General*, 132, 400-418.

Mayr, U., & Keele, S. W. (2000). Changing internal constraints on action: The role of backward inhibition. *Journal of Experimental Psychology: General*, 129, 4-26.

Mayr, U., Kuhns, D., & Hubbard, J. (2014). Long-term memory and the control of attentional control. *Cognitive Psychology*, 72, 1-26.

Mayr, U., Kuhns, D., & Rieter, M. (2013). Eye movements reveal dynamics of task control. *Journal of Experimental Psychology: General*, 142, 489-509.

Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 22, 1423-1442.

Meiran, N. (2000). Reconfiguration of stimulus task sets and response task sets during task switching. In J. Driver & S. Monsell (Eds.), *Control of Cognitive Processes: Attention and Performance XVIII* (pp. 377-399). Cambridge, MA: MIT Press.

Meiran, N., Chorev, Z., & Sapir, A. (2000). Component processes in task switching. *Cognitive Psychology*, 41, 211-253.

Meiran, N., & Kessler, Y. (2008). The task rule congruency effect in task switching reflects activated long-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 137–157.

Meiran, N., & Marciano, H. (2002). Limitations in advance task preparation: Switching the relevant stimulus dimension in speeded same-different comparisons. *Memory and Cognition*, 30, 540-550.

Meiran, N., Kessler, Y., & Adi-Japha, E. (2008). Control by action representation and input selection (CARIS): A theoretical framework for task switching. *Psychological Research*, 72, 473-500.

Meyer, D. E., Osman, A. M., Irvin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology*, 26, 3-67.

Miller, J., Patterson, T., & Ulrich, R. (1998). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, 35, 99-115.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, 7, 134-140.

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., & Mizon, G. A. (2006). Can the task-cueing paradigm measure "endogenous" task-set reconfiguration? *Journal of Experimental Psychology: Human Perception and Performance*, 32, 493-516.

Monsell, S., Sumner, P., & Waters, H. (2003). Task-set reconfiguration with predictable and unpredictable task switches. *Memory and Cognition*, 31, 327–342.

Monsell, S., Taylor, T. J., & Murphy, K. (2001). Naming the color of a word: Is it responses or task sets that compete? *Memory and Cognition*, 29, 137-151.

Müller, H. J., Reimann, B., & Krummenacher, J. (2003). Visual search for singleton feature targets across dimensions: Stimulus- and expectancy-driven effects in dimensional weighting. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 1021-1035.

Nako, R., Wu, R., & Eimer, M. (2014). Rapid guidance of visual search by object categories. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 50-60.

Nicholson, R., Karayanidis, F., Poboka, D., Heathcote, A., & Michie, P.T. (2005). Electrophysiological correlates of anticipatory task-switching processes. *Psychophysiology*, 42, 540–554.

Nieuwenhuis, S., & Monsell, S. (2002). Residual costs in task switching: Testing the failure to-engage hypothesis. *Psychonomic Bulletin and Review*, 9, 86-92.

Nieuwenhuis, S., Yeung, N., & Cohen, J.D. (2004). Stimulus modality, perceptual overlap, and the Go/NoGo N2. *Psychophysiology*, 41, 157-160.

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, 113, 327-357.

Pashler, H. (2000). Task switching and multitask performance. In Monsell, S., & Driver, J. (Eds.). *Control of mental processes: Attention and Performance* XVIII (pp. 277-307). Cambridge, MA: MIT Press.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3-25.

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.

Proverbio, A. M., Vecchi, L., & Zani, A. (2004). From orthography to phonetics: ERP Measures of grapheme-to-phoneme conversion mechanisms in reading. *Journal of Cognitive Neuroscience*, 16, 301-317.

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

Rogers, R. D., & Monsell, S. (1995). The costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124, 207-231.

Rubinstein, J. S., Meyer, D. E., & Evans, E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 763 – 797.

Rushworth, M. F. S., Passingham, R. E., & Nobre, A. C. (2002). Components of switching intentional set. *Journal of Cognitive Neuroscience*, 14, 1139-1150.

Rushworth, M. F. S., Passingham, R. E., & Nobre, A. C. (2005). Components of attentional set switching. *Experimental Psychology*, 52, 83-98.

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.

Schneider, D. W., & Logan, G. D. (2005). Modeling task switching without switching tasks: A short-term priming account of explicitly cued performance. *Journal of Experimental Psychology: General*, 134, 343-367.

Schuch, S., & Koch, I. (2003). The role of response selection for inhibition of tasks sets in task shifting. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 92-105.

Sereno, S.C., Rayner, K., & Posner, M.I. (1998). Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *NeuroReport*, 9, 2195-2200.

Spence, C., & Driver, J. (1997). On measuring selective attention to an expected sensory modality. *Perception & Psychophysics*, 59, 389-403.

Steinhauser, M., & Hübner, R. (2007). Automatic activation of task-related representations in task shifting. *Memory and Cognition*, 35, 138-155.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, 135, 77-99.

Tieges, Z., Snel, J., Kok, A., Wijnen, J. G., Lorist, M. M., & Ridderinkhof, K. R. (2006). Caffeine improves anticipatory processes in task switching. *Biological Psychology*, 73, 101–113.

Tieges, Z., Snel, J., Kok, A., Plat, N. & Ridderinkhof, K. R. (2007). Effects of caffeine on anticipatory control processes: Evidence from a cued task-switching paradigm. *Psychophysiology*, 44, 561–578.

Treisman, A. (1969). Strategies and models of selective attention. *Psychological Review*, 76, 282-299.

Treisman, A. M. (1971). Shifting attention between the ears. *Quarterly Journal of Experimental Psychology*, 23, 157-167.

Vachon, F., & Jolicoeur, P. (2011). Impaired semantic processing during task-set switching: Evidence from the N400 in rapid serial visual presentation. *Psychophysiology*, 48, 102-111.

Vandierendonck, A., Liefooghe, B., & Verbruggen, F. (2010). Task Switching: interplay of reconfiguration and interference control. *Psychological Bulletin*, 136, 601-626.

Waszak, F., Hommel, B., & Allport, A. (2003). Task-switching and long-term priming: Role of episodic S-R bindings in task-shift costs. *Cognitive Psychology*, 46, 361-413.

Waszak, F., Hommel, B., & Allport, A. (2004). Semantic generalization of stimulus-task bindings. *Psychonomic Bulletin and Review*, 11, 1027-1033.
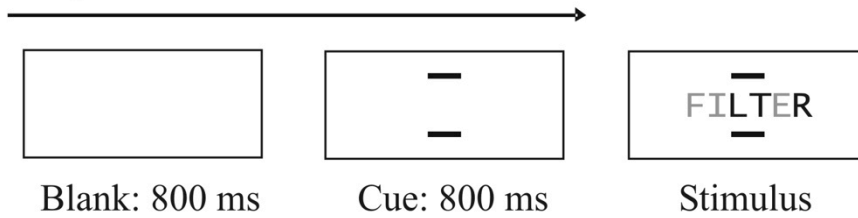
Waszak, F., Hommel, B., & Allport, A. (2005). Interaction of task readiness and automatic retrieval in task switching: Negative priming and competitor priming. *Memory and Cognition*, 33, 595–610

Wendt, M., & Kiesel, A. (2008). The impact of stimulus-specific practice and task instructions on response congruency effects between tasks. *Psychological Research*, 72, 425-432.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5, 495-501.

Yeung, N., & Monsell, S. (2003). Switching between tasks of unequal familiarity: The role of stimulus-attribute and response-set selection. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 455-469.

Yeung, N., Nystrom, L. E., Aronson, J. A., & Cohen, J. D. (2006). Between-task competition and cognitive control in task switching. *Journal of Neuroscience*, 26, 1429-1438.

FIGURES AND LEGENDS

A. a trial

Response to Stimulus Interval 1600 ms

Blank: 800 ms   Cue: 800 ms   Stimulus

B. Cues with examples of symmetric and asymmetric stimuli

Figure 1. Experiment 1: trial structure, cues and stimuli; grey and black letters stand for blue and red letters in the actual displays.
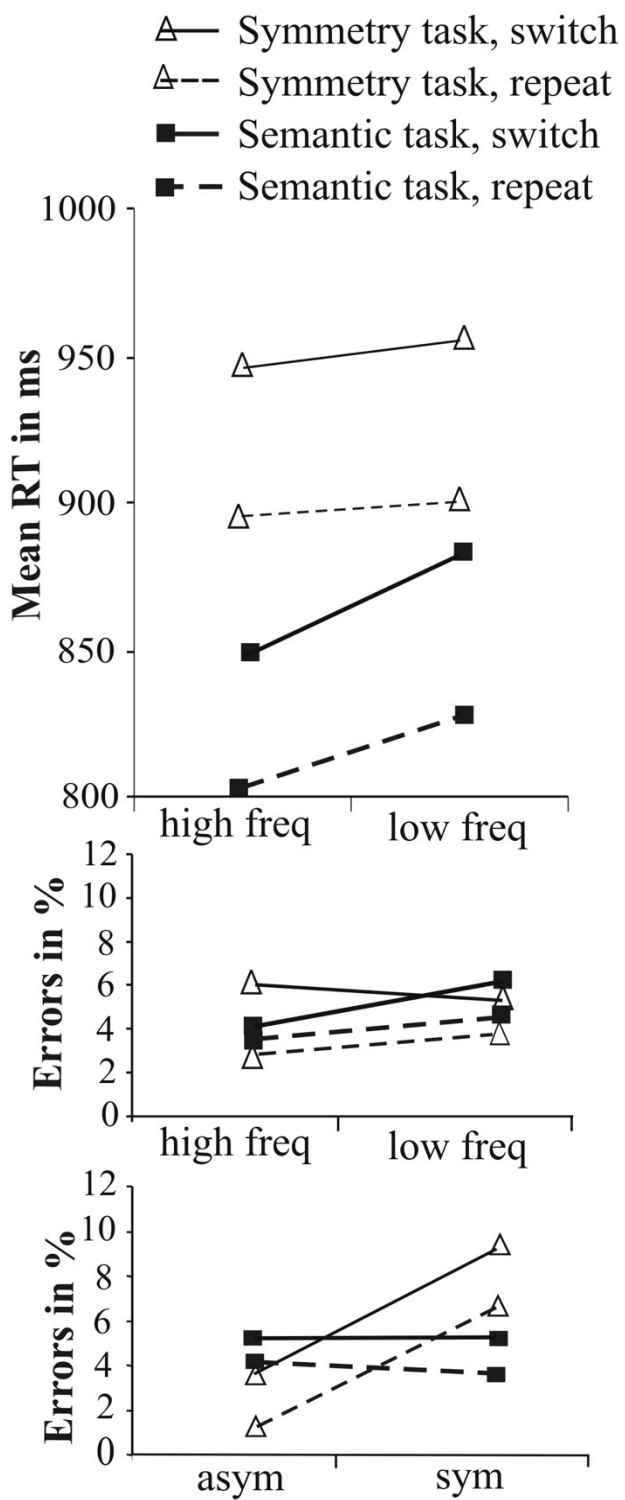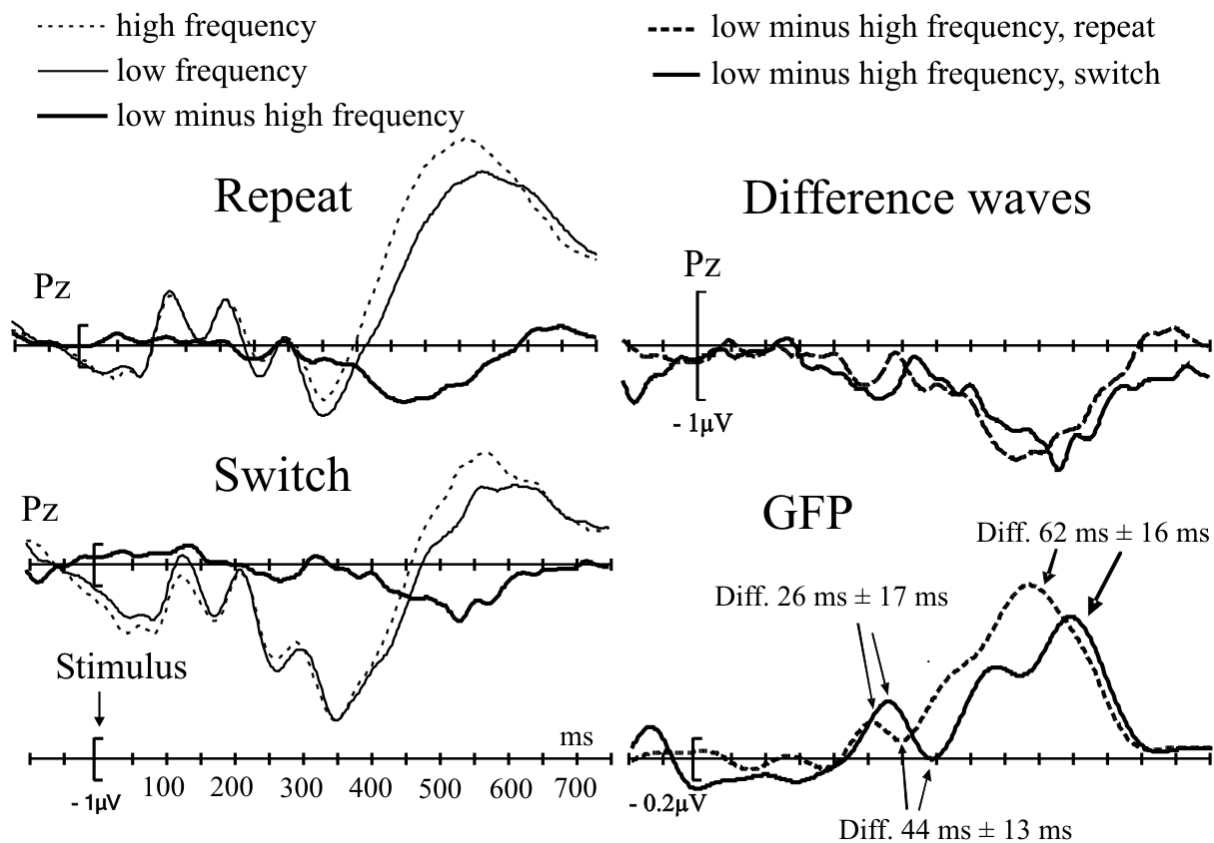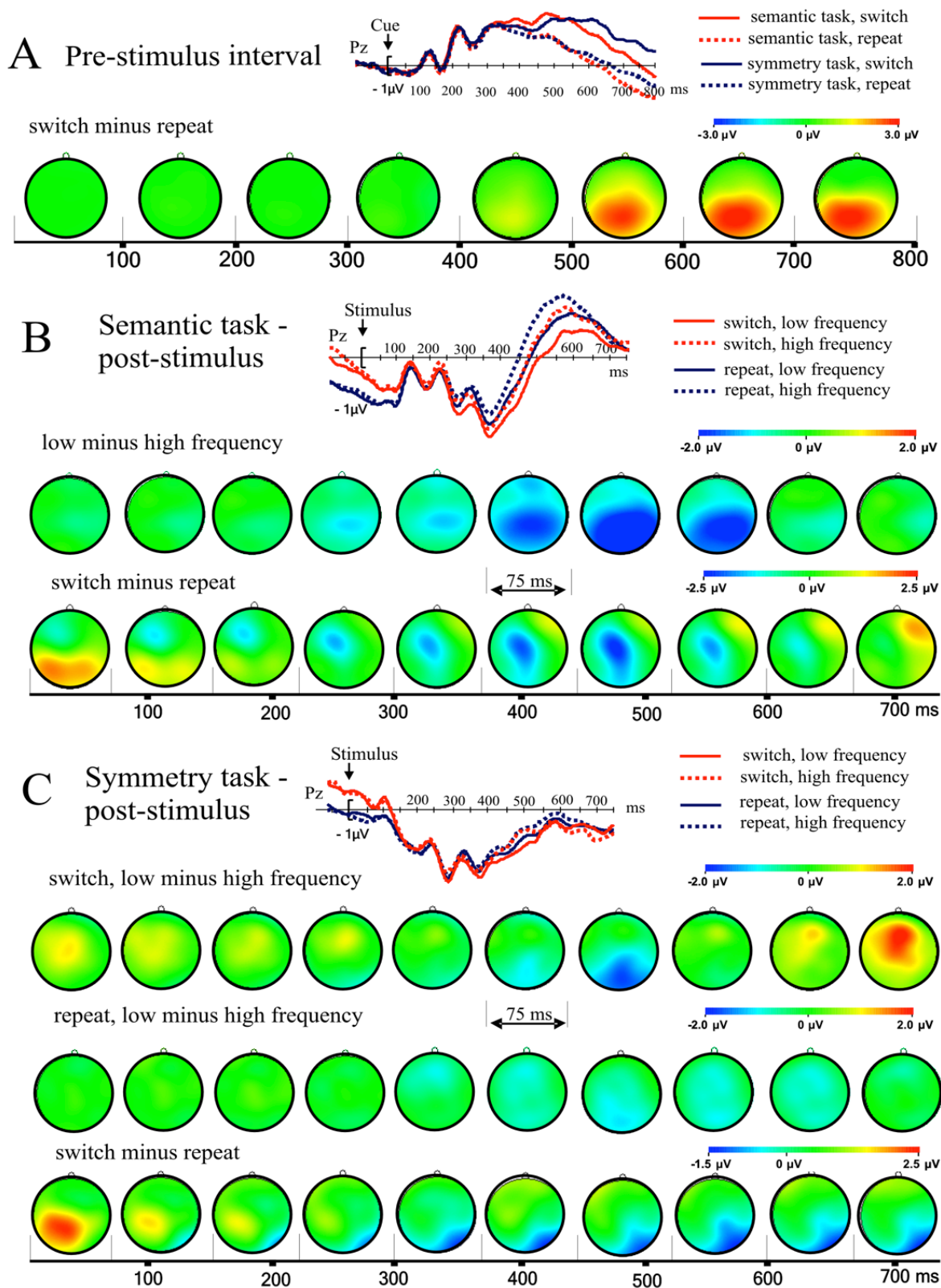
Figure 2. Behavioral results for Experiment 1.
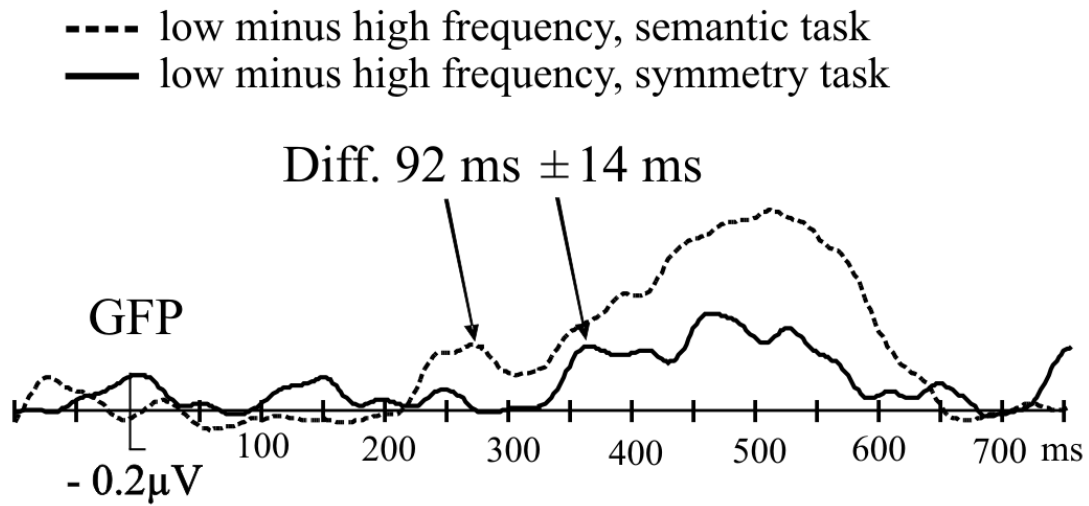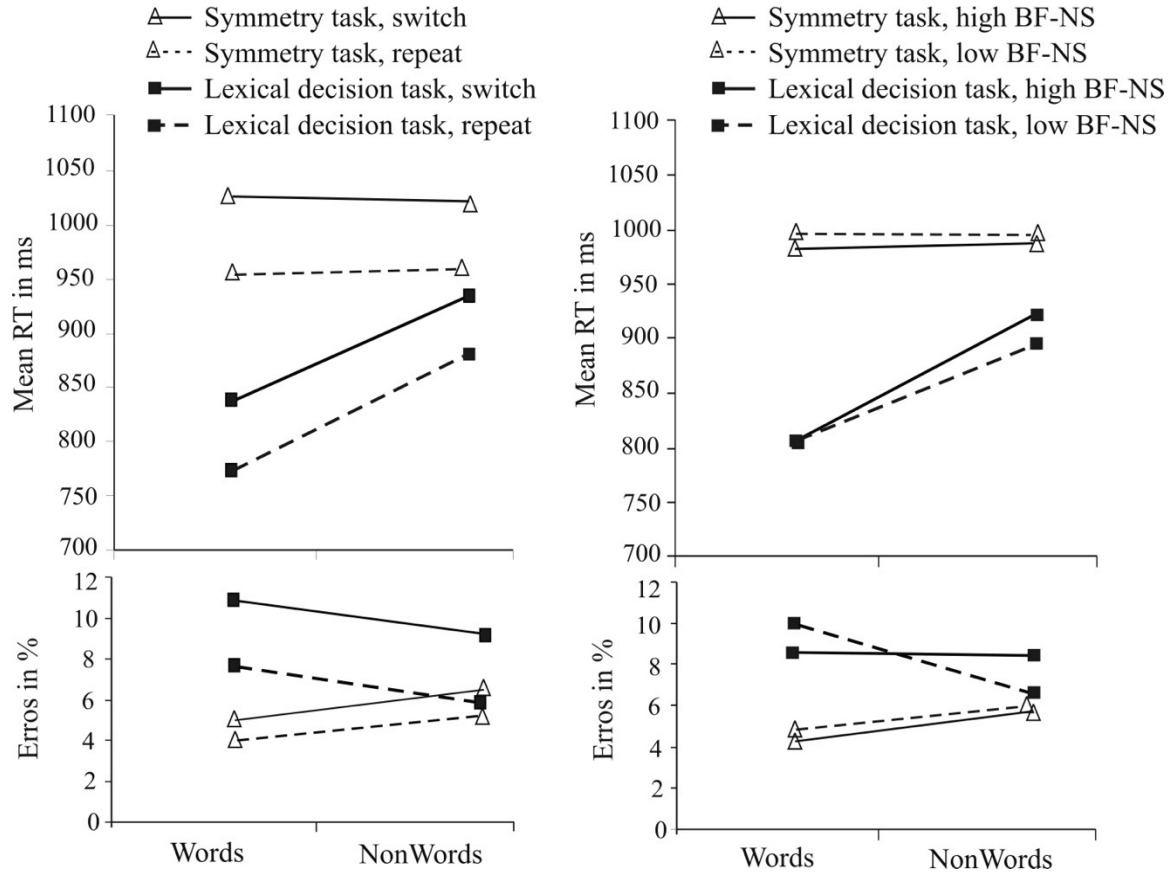
Figure 3. Experiment 1: electrophysiological data and results from the latency analysis. Left

panels: ERPs in a sample electrode, high frequency-minus-low frequency. Right panels:

difference waves in the same electrode and the Global Field Power (GFP) – amplitude measure

pooled over all scalp electrodes – of the frequency difference with estimated latency differences.

A Pre-stimulus interval

semantic task, switch
semantic task, repeat
symmetry task, switch
symmetry task, repeat

switch minus repeat

-3.0 µV   0 µV   3.0 µV

100   200   300   400   500   600   700   800

B Semantic task - post-stimulus

switch, low frequency
switch, high frequency
repeat, low frequency
repeat, high frequency

low minus high frequency

-2.0 µV   0 µV   2.0 µV

switch minus repeat

75 ms

-2.5 µV   0 µV   2.5 µV

100   200   300   400   500   600   700 ms

C Symmetry task - post-stimulus

switch, low frequency
switch, high frequency
repeat, low frequency
repeat, high frequency

switch, low minus high frequency

-2.0 µV   0 µV   2.0 µV

repeat, low minus high frequency

75 ms

-2.0 µV   0 µV   2.0 µV

switch minus repeat

-1.5 µV   0 µV   2.5 µV

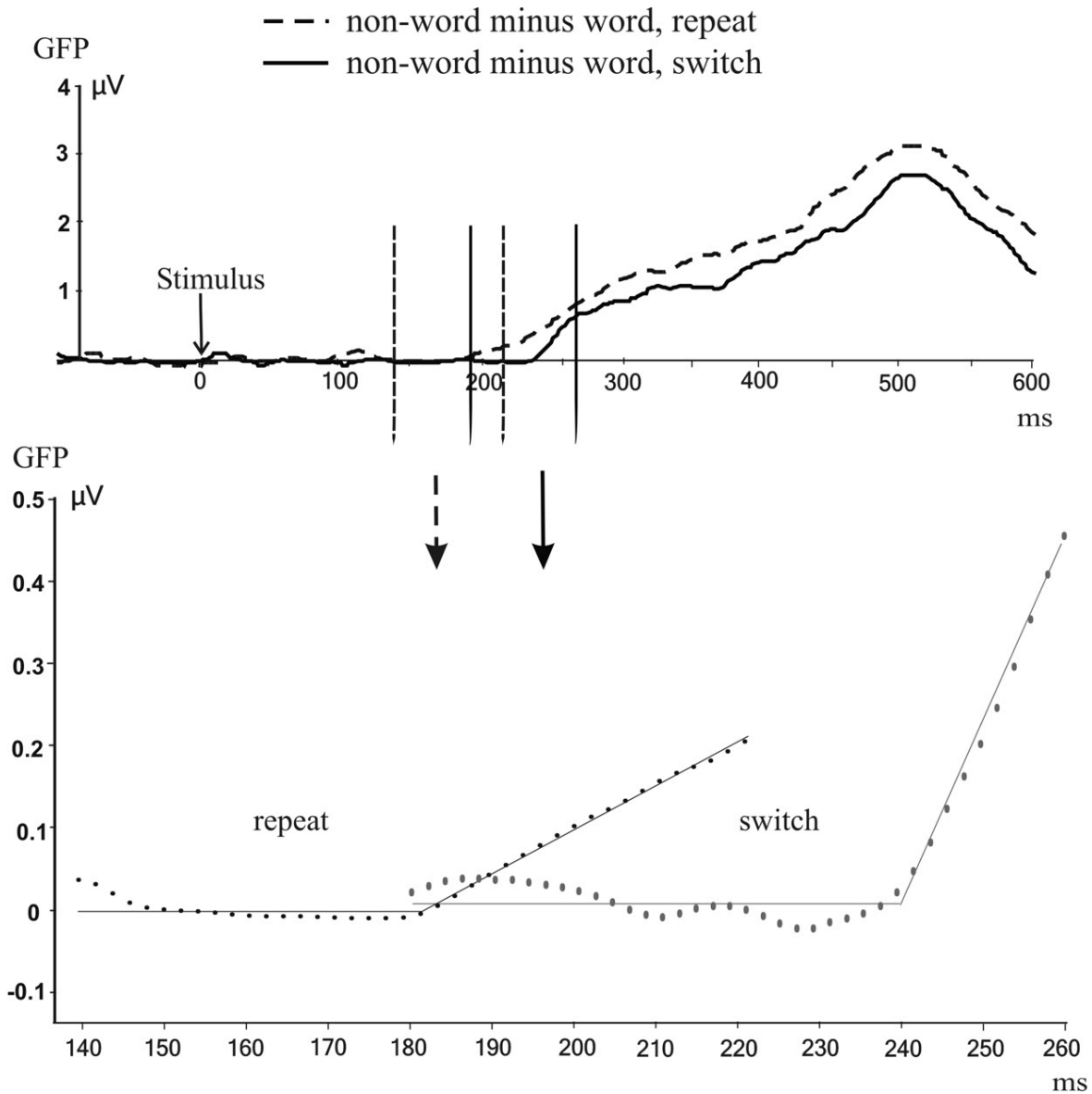100   200   300   400   500   600   700 ms

Figure 4. Experiment 1: ERP amplitude differences. **A:** Pre-stimulus ERPs for the semantic and symmetry tasks in Pz; scalp distributions (topographies) of the switch-repeat difference averaged across each 100 ms window up to stimulus onset. For semantic task trials (**B**) and symmetry task trials (**C**): topographies of the effects of a task-switch and of lexical frequency averaged across each 75-ms window following stimulus onset (with illustrative waveforms for electrode Pz).

Figure 5. Experiment 1: Latency analysis of the Global Field Power (GFP) of the frequency effect in the semantic and symmetry tasks.
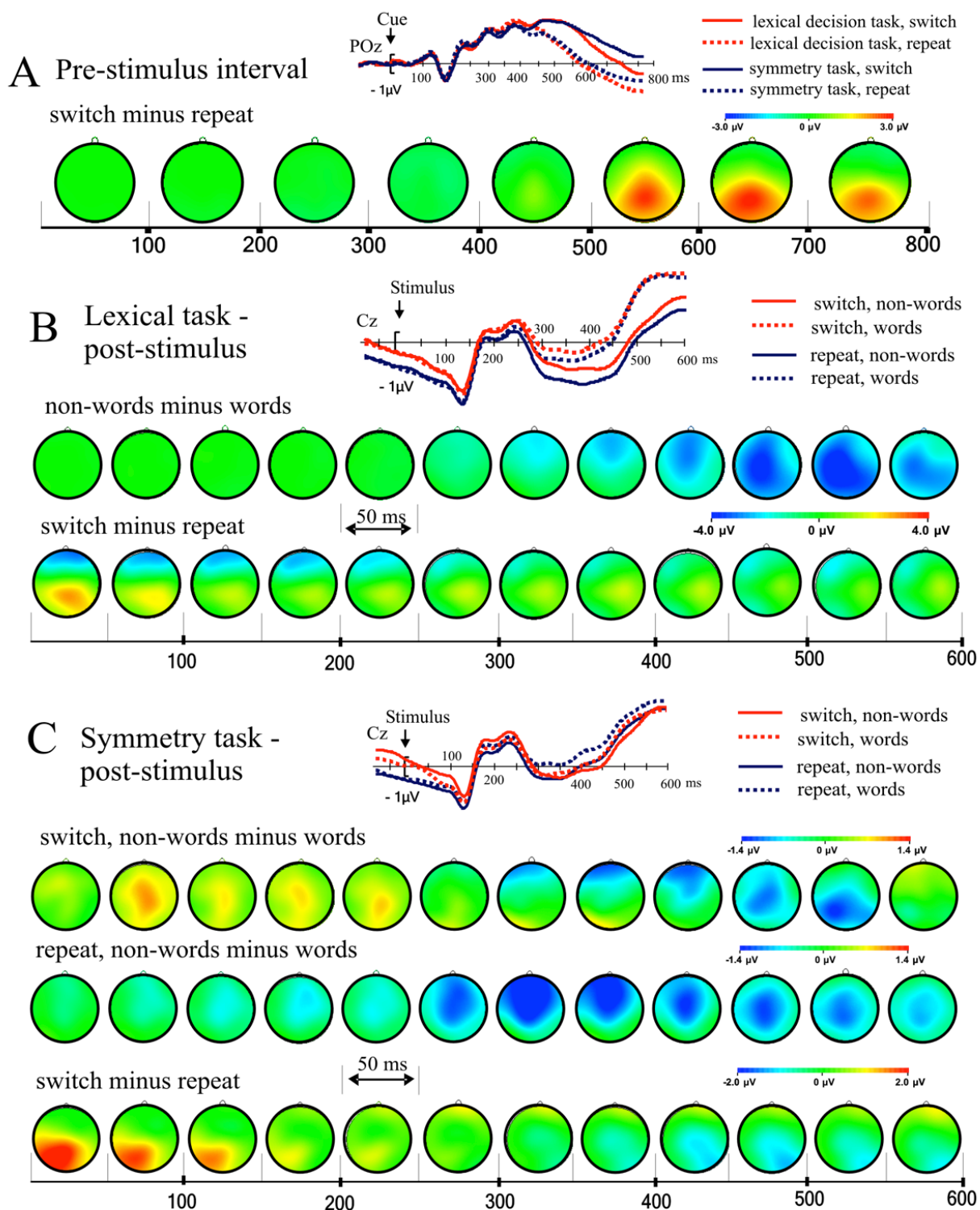
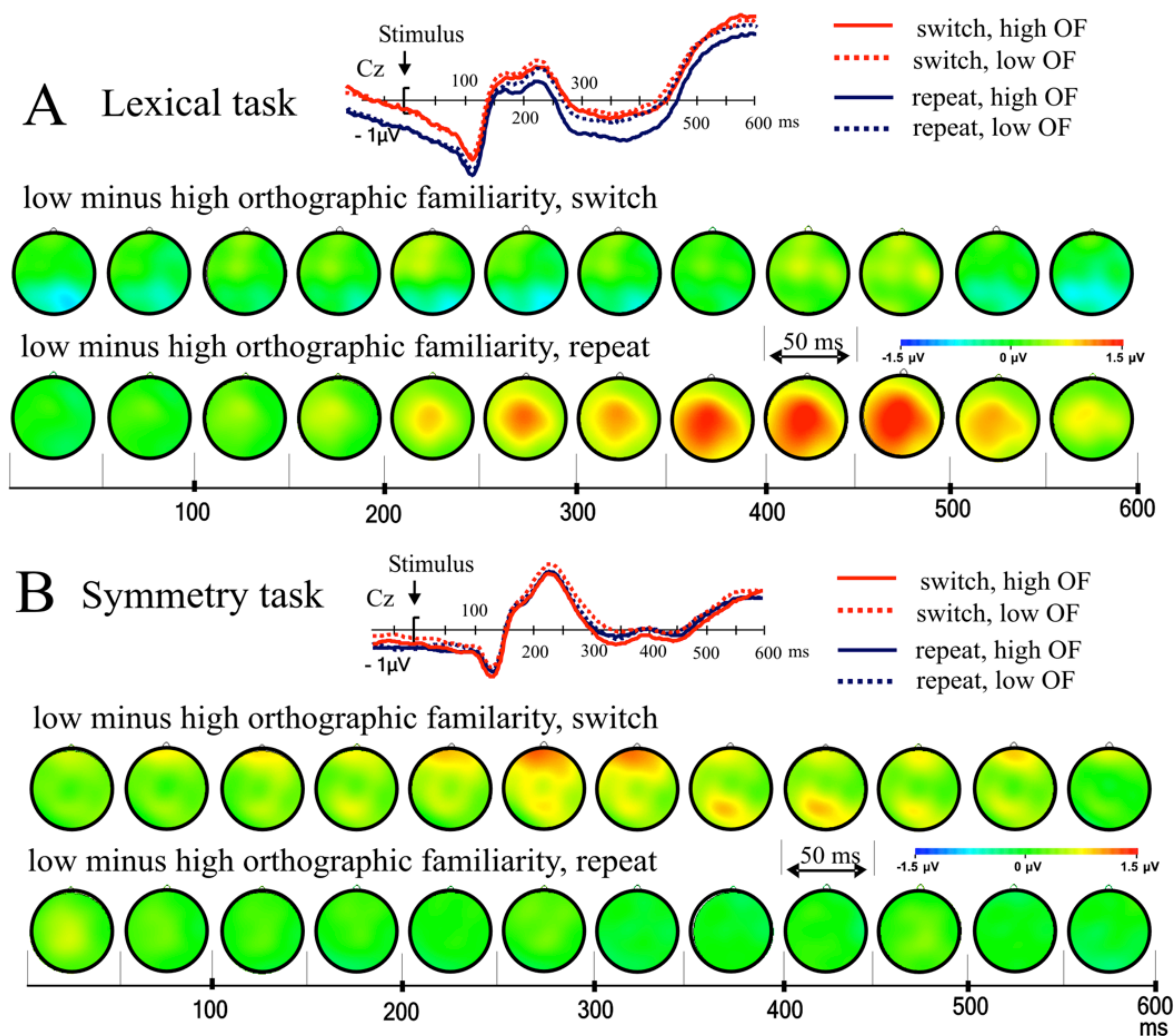**Figure 6.** Behavioral results for Experiment 2.

Figure 7. Experiment 2: Latency analysis of the lexicality effect (computed as Global Field Power of the nonword-word difference waves). The vertical lines in the top panel indicate the stretches of the data subjected to bilinear function fitting (broken lines for repeats and solid lines for switches). The lower panel zooms in on these time ranges.

Figure 8. Experiment 2: ERP amplitude differences. **A:** Pre-stimulus ERPs for the lexical

decision and symmetry tasks in POz and topographies of the switch-repeat difference averaged

across each 100 ms window up to stimulus onset. For lexical decision trials (**B**) and symmetry

task trials (**C**): topographies of the effects of a task-switch and lexicality averaged across each 50

ms window following stimulus onset (with illustrative waveforms for electrode Cz).

Figure 9. Experiment 2: Topographies of the effect of low and high orthographic familiarity for task switch and repeat trials, averaged across 50-ms time-windows following stimulus onset, with illustrative waveforms from electrode Cz, for lexical decision (A) and the symmetry task (B).

Figure 10. Global Field Power (GFP) of the lexicality effect in the lexical and symmetry task, illustrating the effects of lexicality on ERP latency and amplitude.

TABLES

	Semantic task				Symmetry task			
	Frequency		Symmetry		Frequency		Symmetry	
	High	Low	Asym	Sym	High	Low	Asym (Go)	Sym (Nogo)
Switch	848 (4.1)	883 (6.2)	854 (5.2)	885 (5.3)	946 (6)	956 (5.3)	951 (3.6)	– (9.2)
Repeat	803 (3.5)	828 (4.5)	813 (4.1)	819 (3.6)	896 (2.9)	901 (3.8)	899 (1.2)	– (6.6)
Switch cost	45 (0.6)	55 (1.7)	42 (1.1)	65 (1.7)	50 (3)	54 (1.6)	52 (4)	– (2.7)

Table 1. Experiment 1, mean RTs in ms, error rates in % and RT and error switch costs.

	Lexical task				Symmetry task			
	Lexicality		Othogr. Fam.		Lexicality		Othogr. Fam.	
	NonW	Word	High	Low	NonW	Word	High	Low
Switch	934 (9.2)	838 (10.9)	893 (10.0)	879 (10.1)	1022 (6.5)	1025 (5.0)	1018 (5.5)	1030 (6.0)
Repeat	880 (5.7)	774 (7.6)	833 (7.0)	820 (6.4)	959 (5.2)	953 (4.0)	952 (4.4)	960 (4.8)
Switch cost	54 (3.5)	64 (3.3)	60 (3.0)	59 (3.7)	63 (1.3)	73 (1.0)	66 (1.1)	70 (1.2)

Table 2. Experiment 2, mean RTs in ms, error rates in % and switch costs by task, lexicality and orthographic familiarity.

APPENDIX

Conditions	dfs	F	p
Switch	1,17	34.57	< .001
switch x region	3,51	19.04	< .001
switch x laterality	2,34	6.7	< .01
switch x region x laterality	6,102	4.31	< .01
switch x task x laterality	2,34	6.2	< .01

Table A1. Experiment 1, significant effects and interactions in pre-stimulus ERP amplitudes.

Time window	Conditions	dfs	F	p
100 – 200 ms	switch x laterality	2,34	4.27	< .05
	switch x region x laterality	6,102	4.56	< .05
200 – 300 ms	frequency	1,17	5.66	< .05
	switch x laterality	2,34	8.18	< .01
	switch x region x laterality	6,102	4.24	< .05
300 – 400 ms	frequency	1,17	7.85	< .05
	switch x laterality	2,34	16.47	< .001
	switch x region x laterality	6,102	4.06	< .05
400 – 500 ms	frequency	1,17	5.56	< .05
	switch	1,17	16.49	< .001
	switch x laterality	2,34	4.52	< .05

Table A2. Significant effects and interactions in post-stimulus ERP amplitudes in the semantic task.

Time window	Conditions	dfs	F	p
100 – 200 ms	switch	1,17	7.37	< .05
	switch x laterality	2,34	6.85	< .01
200 – 300 ms	switch x laterality	2,34	5.86	< .01
	frequency x region x laterality	6,102	2.36	< .05
400 – 500 ms	frequency	1,17	7.31	< .05
	switch x region	3,51	4.61	< .05
	switch x laterality	2,34	3.56	< .05
	frequency x region	3,51	8.03	< .01

Table A3. Significant effects and interactions in post-stimulus ERP amplitudes in the symmetry task.

	Words- Low	Words- High	NonWords-Low	NonWords-High
4 letters	28	26	25	19
5 letters	115	109	109	102
6 letters	189	185	187	190
7 letters	268	280	279	289
1 syllable	115	117	105	79
2 syllables	368	275	386	479
3 syllables	102	100	98	45
4 syllables	15	8	12	0
2 phonemes	4	0	0	0
3 phonemes	39	33	11	12
4 phonemes	120	133	77	104
5 phonemes	186	203	197	158
6 phonemes	172	164	223	223
7 phonemes	71	64	92	103
8 phonemes	8	3	0	0

Table A4. Numbers of items presented for different conditions (cells).

	Lexical task		Symmetry task	
	Congruent	Incongruent	Congruent	Incongruent
Switch	881 (4.7)	891 (15.4)	1010 (3.3)	1037 (8.2)
Repeat	822 (4.9)	831 (8.5)	954 (3.7)	958 (5.5)
Switch cost	59 (-0.2)	60 (6.9)	57 (-0.4)	79 (2.7)

Table A5. Experiment 2, mean RTs in ms, error rates in % and switch costs by task and congruency.

Conditions	dfs	F	p
switch	1,17	35.04	< .001
switch x region	3,51	35.47	< .001
switch x laterality	2,34	16.19	< .001
switch x region x laterality	6,102	3.86	< .01
switch x task x laterality	2,34	5.49	< .05

Table A6. Experiment 2, significant effects and interactions in the pre-stimulus ERP amplitudes.

Time window	Conditions	dfs	F	p
100 – 200 ms	switch x region	3,51	17.32	< .001
	OF x laterality	2,34	4.15	< .05
200 – 300 ms	lexicality	1,17	11.89	< .01
	lexicality x region	3,51	4.88	< .05
	lexicality x laterality	2,34	4.28	< .05
	switch x region	3,51	10.78	< .001
	OF x region	3,51	5.13	< .05
	OF x laterality	2,34	5.58	< .05
	switch x OF	1,17	5.44	< .05
	repeat only:			
	OF	1,17	8.29	< .05
	switch only:			
	OF x region	3,51	7.31	< .01
300 – 400 ms	lexicality	1,17	30.18	< .001
	lexicality x region	3,51	9.69	< .001
	lexicality x region x laterality	6,102	2.74	< .05
	switch x region	3,51	4.21	< .05
	switch x laterality	2,34	4.89	< .05
	OF x laterality	2,34	3.62	< .05
	switch x OF	1,17	8.81	< .01
	repeat only:			
	OF	1,17	13.34	< .01

400 – 500 ms	lexicality	1,17	29.8	< .001
	lexicality x laterality	2,34	9.13	< .001
	lexicality x region x laterality	6,102	5.46	< .001
	switch x laterality	2,34	6.35	< .05
	OF	1,17	8.74	< .01
	OF x laterality	2,34	4.91	< .05
	switch x OF	1,17	4.68	< .05
	switch x OF x region x laterality	6,102	2.25	< .05
	repeat only:			
	OF	1,17	14.37	< .001
	OF x laterality	2,34	7.07	< .01
	OF x region x laterality	6,102	2.71	< .05

Table A7. Significant effects and interactions of post-stimulus ERP amplitudes in the lexical task (OF = orthographic familiarity).

Time window	Conditions	dfs	F	p
100 – 200 ms	switch x laterality	2,34	5.49	< .05
	switch x region x laterality	6,102	4.98	< .001
	switch x lexicality	1,17	9.3	< .001
	switch x lexicality x laterality	2,34	7.19	< .001
	switch only:			
	lexicality	1,17	6.18	< .05
	repeat only:			
	lexicality x laterality	2,34	5,12	< .05
200 – 300 ms	switch x region x laterality	6,102	3.69	< .01
	switch x lexicality	1,17	6.25	< .05
	switch x lexicality x laterality	2,34	5.27	< .05
	switch only:			
	OF	1,17	4.93	< .05
	repeat only:			
	lexicality	1,17	8.35	< .05
	lexicality x laterality	2,34	4.03	< .05
300 – 400 ms	lexicality	1,17	7.93	< .05
	lexicality x region	3,51	15.93	< .001
	lexicality x region x laterality	6,102	5.6	< .001
	switch x region x laterality	6,102	4.16	< .01
400 – 500 ms	switch	1,17	6.01	< .05
	switch x region	3,51	7.51	< .01
	switch x laterality	2,34	4.43	< .05

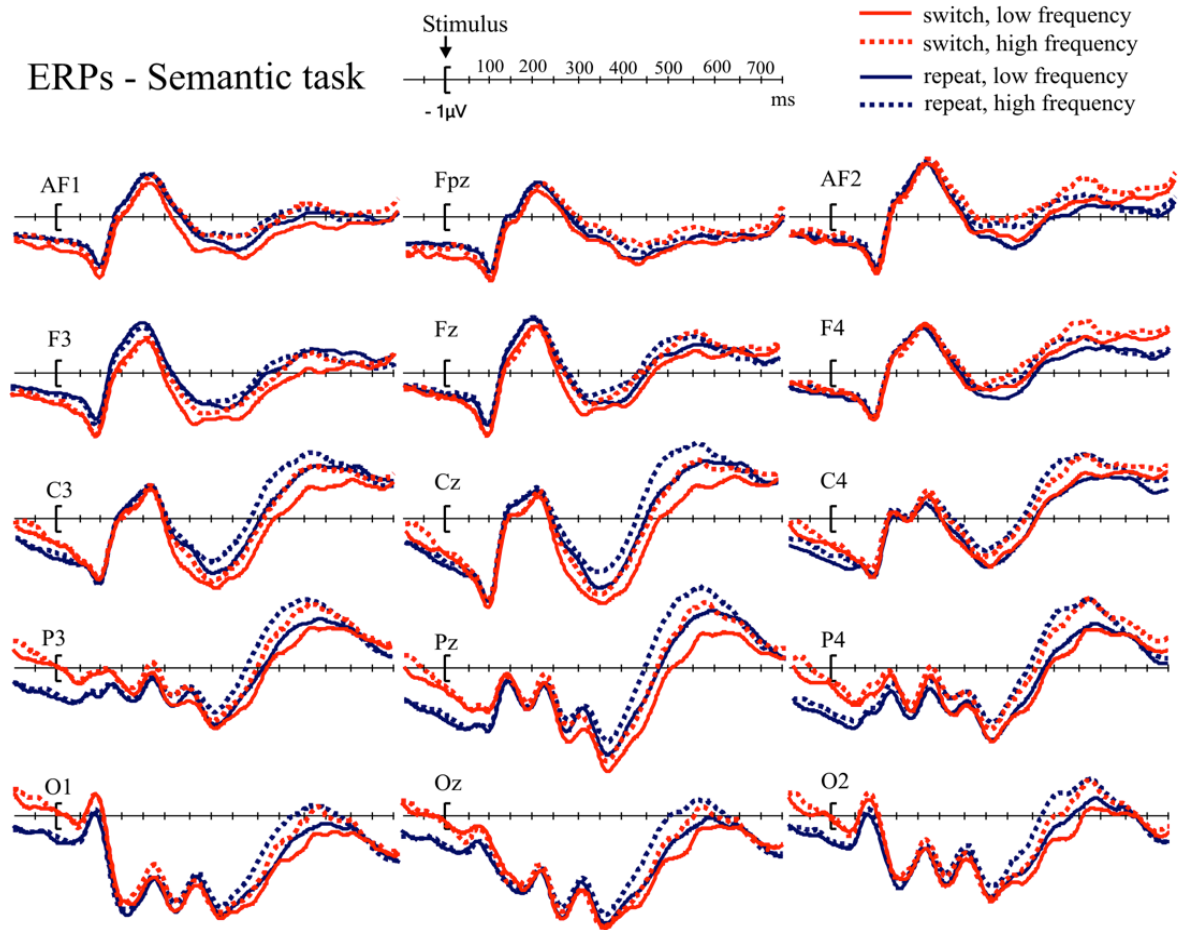|			
switch x region x laterality	6,102	5.78	< .001
lexicality x region x laterality	6,102	3.4	< .05

Table A8. Significant effects and interactions in the post-stimulus ERP amplitudes in the symmetry task.
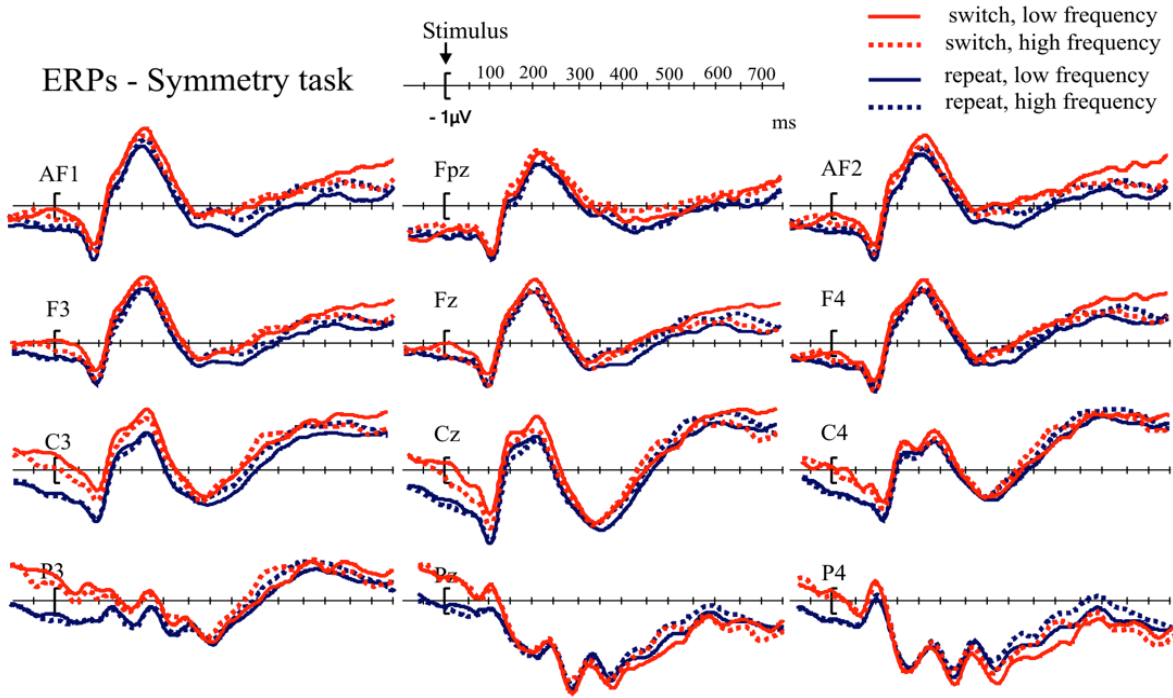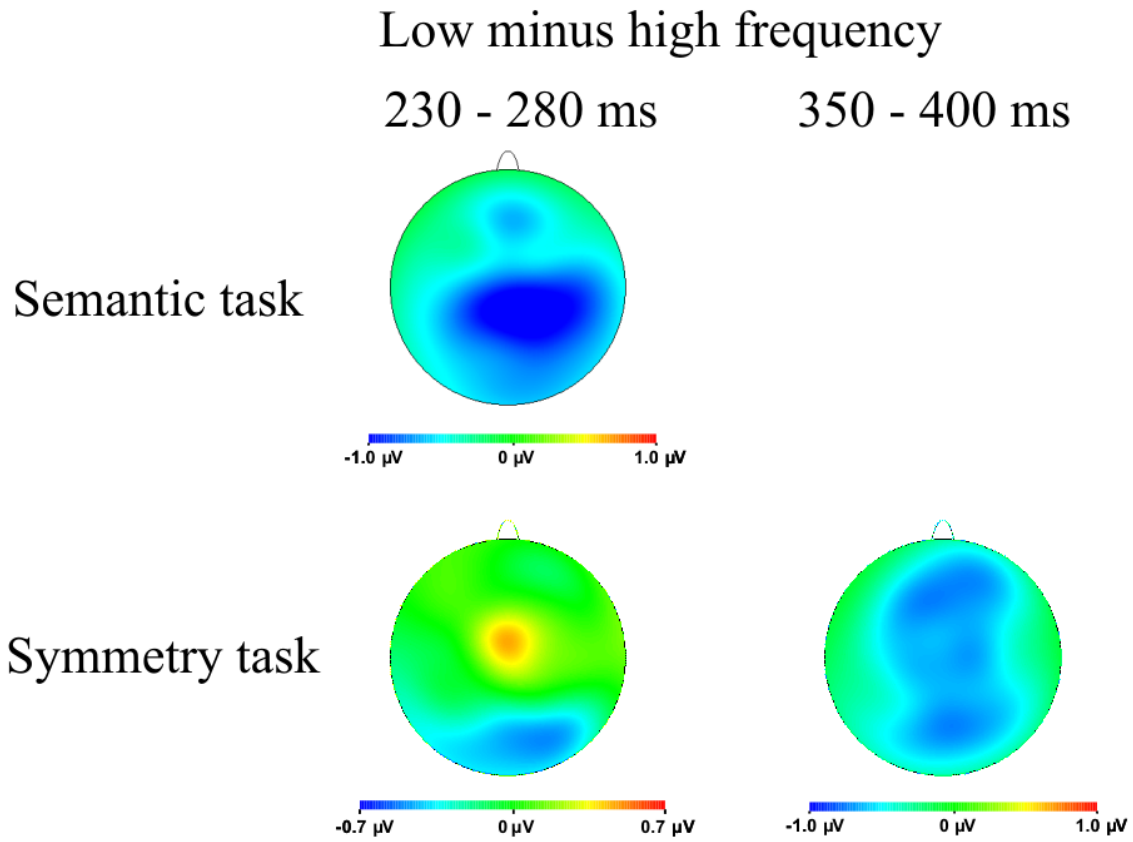
SUPPLEMENTARY MATERIALS

Figure S1. Grouping of electrodes into 12 regions by the anterior-posterior (4 levels) and laterality (3 levels) dimensions.
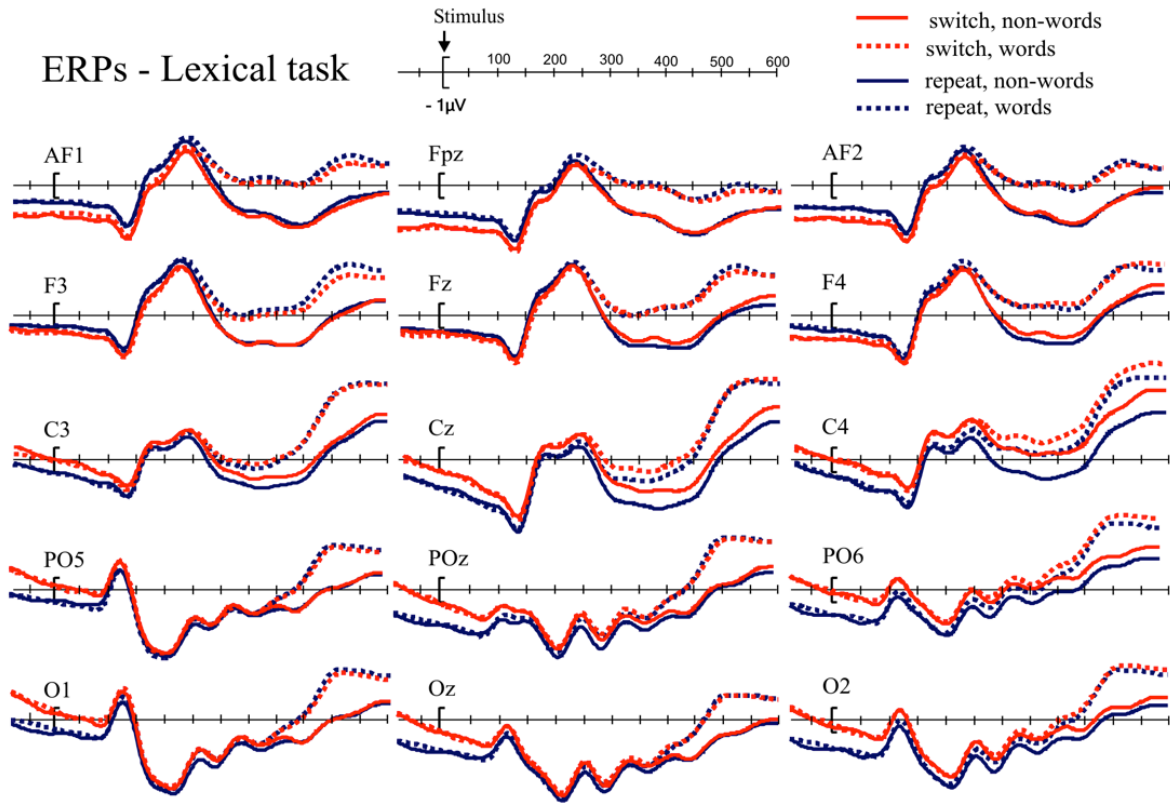
Figure S2. Experiment 1, semantic task: low and high frequency ERP waveforms for switch and repeat trials in selected electrodes.
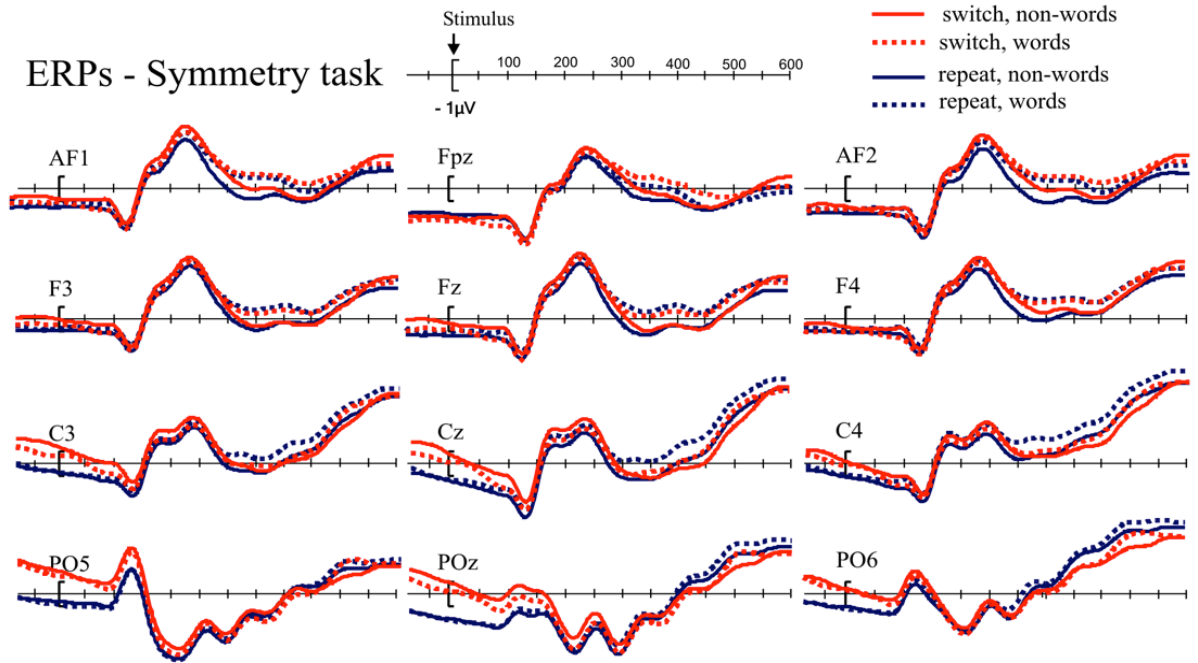
Figure S3. Experiment 1: symmetry task, low and high frequency ERP waveforms for switch and repeat trials in selected electrodes**.**

Figure S4. Experiment 1: scalp distribution of frequency effect by task (pooling over switch and repeat) at intervals around critical peaks in the GFP of the difference waves in Figure 5.

Figure S5. Experiment 2, lexical decision task: ERP waveforms for non-words and words, for switch and repeat trials, in selected electrodes.

Figure S6. Experiment 2, symmetry task: ERP waveforms for non-words and words, for switch and repeat trials, in selected electrodes.

Changes made in final revision – indexed to Editor comments:

p. 15 - Please move the lexicality definition from footnote 2 to the main text.

First full para on p.15 rewritten to incorporate the definition of "lexicality" in the text and clarify the logic.

p. 27 - Please make it clearer that the topographic maps are averages across each 75 ms window.

Captions of Fig 4 and 8 rewritten to clarify the averaging interval for each topographic map series.

p. 38 - APA copyeditors will not like your citing of Figure 9 before Figure 8 in the text, so you should make the text citation order consistent with the figure numbering. One possibility is to move (something like) this whole paragraph from p 38 to p 39 where the main discussion of Figure 9 appears.

We have done what you suggest, removing the paragraph on p,38, and incorporating the point in first full para on p.39 -- this also eliminates awkward back and forth references and abbreviates the exposition.

p. 41 - This is a critical sentence, but very difficult to understand as written: "ERP waveforms for low and high frequency words started to diverge at ~250 ms and switching reliably delayed this onset by a substantial proportion of the RT switch cost as early as ~300 ms. " The following summary of Experiment 2 seems much better, so I would suggest changing to something more like this format for summarizing Experiment 1: "The onset of the lexicality effect at ~200 ms was delayed on switch trials by about 60 ms, again an amount similar to the switch cost. "

We see the problem, and have tried to clarify. The expression cannot be quite parallel for the two experiments as in Exp 1, as we don't really see a delay in onset of the frequency effect on switch trials, rather a delay in what happens to it thereafter

Figure 3 caption - Include "Global Field Power (GPF)" to define GFP abbreviation in figure.

Caption now says: "and the Global Field Power (GFP) – amplitude measure pooled over all scalp electrodes – of the frequency difference"

Figure S3 caption - Change "trals" to "trials"

Oops. "I" inserted.

Other minor changes:

p.14 "the latter finding" instead of "this finding" to clarify which of Hsieh et al's findings we are talking about.
p.37 "typical of" changed to "as expected for"