Available online at www.sciencedirect.com

SCIENCE DIRECT®

International Journal of Psychophysiology 60 (2006) 240 – 250

INTERNATIONAL JOURNAL OF PSYCHOPHYSIOLOGY

www.elsevier.com/locate/ijpsycho

P300-based Stroop study with low probability and target Stroop oddballs: The evidence still favors the response selection hypothesis

J. Peter Rosenfeld *, Katieann R. Skogsberg

Psychology Department, Northwestern University, Evanston, IL 60208, United States

Received 16 November 2004; received in revised form 25 March 2005; accepted 31 May 2005
Available online 15 August 2005

Abstract

This paper addresses the issue of the locus of action in cognitive processing during Stroop effects. It uses the P300 latency to assess stimulus processing effects, but, for the first time, under conditions in which Stroop stimuli are rare and target stimuli. The study was also concerned with demonstrating that apparent P300s during verbal responding are in fact uninterpretable due to contamination of EEG by speech-related artifact. Three studies were presented. In Study 1, there were 3 blocks, each containing 1 of 3 types of rare Stroop stimuli ($p = .15$), congruent, neutral, and incongruent. There were also 3 response modes: button press (BUTTON), speaking aloud (VERBAL), and speaking to self (SILENT). Three sessions were used, each for a different response style. The only task was to name the color on each trial. In the 2 non-verbal blocks, Reaction Time (RT) varied by stimulus type; congruent < neutral < incongruent. P300 latency was the same across blocks in these non-verbal conditions in which one saw the classic Pz > Cz > Fz distribution. The much larger, speech artifact-contaminated "P300s" in the VERBAL blocks did suggest a Stroop effect, especially at Fz and Cz, where "P300s" were larger than at Pz. In Study 2, there were 2 response modes, VERBAL and SILENT, and only two rare Stroop stimuli; neutral and incongruent, 1 per block. In each of these blocks, one word–color combination was a designated target requiring a unique response. The subject was to name the color followed by a yes or no to categorize the target or non-target. Again the RT for incongruents was greater than RT for neutrals, without a parallel effect in P300 latency. Again, the rostral ERPs appeared artifactual in the VERBAL condition. Study 3 was a replication of the second study, except that motivated subjects, versus Psychology pool recruits, were used. The latency–RT correlation still failed to obtain. Thus, using classic P300-eliciting antecedents–rare and target (Stroop) stimuli–this study supports the view that the locus of Stroop interference is in response processing.

Keywords: Event-related potential; Stroop task; P300 amplitude and latency; Reaction time

1. Introduction

The Stroop effect (Stroop, 1935) is one of the most robust and best known in psychology. There are actually two Stroop phenomena; Facilitation and Interference. The former effect is demonstrated by comparison of behavioral reaction times (RTs) in response to congruent stimuli (names of colors, e.g., the word, "RED" presented in congruent colors, e.g., red color), in comparison to RTs to neutral stimuli, which are non-color words (e.g., the word, "HAT") presented in some color. The subject's task is to name the color of the displayed word. With congruent stimuli, the facilitated RT is typically less than the neutral RT. On the other hand, in response to incongruent stimuli, such as the word "RED" presented in the color blue, the RT is typically greater than the neutral RT. The typically more robust effect is interference, and the response style yielding the most robust Stroop effects is verbal as opposed to manual responding (MacLeod, 1991).

As discussed by MacLeod (1991) and others, a major theoretical question about the effect has centered on its locus in the cognitive processing sequence. The question is whether Stroop effects are active during the stimulus evaluation phase or the response selection phase or both. Although recent studies have used functional magnetic resonance imaging (fMRI) and other methods so as to find alternative ways to conceptualize the Stroop effect in terms

* Corresponding author.
E-mail address: jp-rosenfeld@northwestern.edu (J.P. Rosenfeld).

0167-8760/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpsycho.2005.05.010

of cognitive control, conflict monitoring and task switching (Kerns et al., 2004; MacDonald et al., 2000), Duncan-Johnson and Kopell (1981) were the first to approach the Stroop mechanism question with a well-conceived psychophysiological study utilizing the latency of the P300 ERP. P300 is an electrically positive-going wave classically elicited during the presentation of a Bernoulli series of rare (e.g., $p<.3$) target stimuli requiring unique behavioral responses and frequent (e.g., $p>.7$) meaningless stimuli; the former but not the latter elicit P300. Since Kutas et al. (1977), Duncan-Johnson (1981), and others had shown that the latency of P300 was a potentially useful index of the stimulus evaluation time, Duncan-Johnson and Kopell (1981) reasoned that if the behavioral Stroop effects were correlated with simultaneously collected P300 latencies, this would support the stimulus evaluation hypothesis of the locus of the Stroop effects. Dissociation of P300 latency and RTs, on the other hand, would support the response selection hypothesis, and indeed, this is what was reported.

There nevertheless were reservations about the strength of support provided by this study due to methodological concerns: Most importantly, the ERP data were collected in a paradigm in which subjects made verbal responses on each trial. It has been long known that speech produces large artifacts in an EEG record (Szirtes and Vaughan, 1977) which precede as well as follow verbalizations by about 1 s. This fact is what initially prompted us (Rosenfeld and Woodley, 1994) to utilize silent (mental) responding, as we will do more systematically, here. Both previously and presently, we observe clear evidence of speech artifact during verbalization trials at all sites. More recently, others have shared this concern and utilized alternative response modes in ERP Stroop studies, e.g., Ilan and Polich (1999), West and Alain (1999), Liotti et al. (2000), who considered all three response modalities which we use here in our Experiment 1, and Atkinson et al. (2003).

Duncan-Johnson and Kopell (1981) moreover presented data from Pz only. The two problem issues associated with this fact are that 1) speech artifact is usually most evident at more anterior sites (as we will also show again here), so that it is not as obvious at Pz; thus noisy Pz data during speech can be mistaken for clean P300 activity; 2) it is impossible without Pz, Fz and Cz data to satisfy the usual scalp distribution attribute that helps identify P300, namely, that Pz>Cz>Fz (Fabiani et al., 1987). If one does not have a definite, artifact-free P300, one cannot confidently utilize the P300 metric of stimulus evaluation in Stroop studies. This latter issue becomes especially important when the classical antecedents for P300 elicitation–rare, target stimuli–are absent, as they were in Duncan-Johnson and Kopell (1981) as well as in the other ERP-Stroop papers cited above. In all these studies, incongruent, neutral, and congruent stimuli were equally probable within a block, and the only task was color naming; i.e., there was no unique target response executed.

The importance of the present study is implied by the preceding two sentences. While it is likely that the late positive components in at least some of the aforementioned, ERP-enhanced Stroop studies based on P300 were or contained elements of the classic P300 or “P3b” components identified by Donchin and Coles (1988) and Fabiani et al. (1987), it is also the case that lacking two of the classic antecedent conditions for P300 elicitation–rare, target stimuli–it could be argued that the late positive components measured were likely to contain other positive components which overlap classic P300 (see Spencer et al., 2001). For example, researchers studying late positive components related to memory attributes usually wish to eliminate confounded probability effects and so use equally probable targets and non-targets. Thus studies using equally probable Stroop stimuli may be looking at memory-related EEG events other than the context updating mechanism theorized by Donchin and Coles (1988) to underlie classic “P3b.” As has been emphasized by Donchin and colleagues many times (though not universally accepted), strict constraints are required for component identification (e.g., Donchin et al., 1978). This is what we have tried to do here–for the first time–that is, utilize the Duncan-Johnson and Kopell (1981) P300 latency–RT correlation approach to investigating the locus of Stroop effects but with particular attention to P300 identification in terms of antecedent conditions emphasized by Donchin and colleagues.

The present set of three experiments was intended to remedy the situation by utilizing rare Stroop stimuli in the first study, and rare, target Stroop stimuli in the latter two studies. Three different response modes are utilized, and ERP data from Pz, Cz, and Fz during speech are compared with those data collected during silent and button-press response blocks. Correlation of P300 Pz latency with behavioral RT, as initially proposed by Duncan-Johnson and Kopell (1981), is the key observation in all studies.

The point of using three modalities here was as follows: The VERBAL modality, in which subjects name colors of stimuli aloud, was used so as to compare our RT and P300 latency data (elicited using classic P300 antecedent conditions) with those of Duncan-Johnson and Kopell (1981), and with special attention to waveforms at anterior sites near oro-facial structures. The manual (button press) and silent responding modalities were used so as to explore results with classic P300-eliciting antecedent conditions; such data did not previously exist.

2. Experiment 1

2.1. Methods

2.1.1. Subjects

There were 14 subjects (six female) run in the study. They were all enrolled in an upper level course in Psychophysiology. All had normal or corrected to normal vision (but no contact lenses). All signed an IRB-approved informed consent form.

2.2. Stimulation and recording procedures

Participants sat in a comfortable padded recliner, and the display face was 1 m from their eyes. The trial structure was as follows: For trial N, EEG recording commenced 104 ms prior to stimulus onset. Recording continued during the 304 ms stimulus exposure, and for the next 1640 ms, yielding a 2048 ms ERP epoch. There followed 900 more ms until the onset of trial $N+1$. Thus the interval between stimuli was about 3 s. Subjects were told to respond as rapidly as possible in all response modalities (see below). EEG was recorded with silver electrodes attached to sites Fz, Cz, and Pz. The scalp electrodes were referenced to linked mastoids. EOG was recorded with silver electrodes above and below the right eye. They were placed intentionally diagonally so they would pick up both vertical and horizontal eye movements as verified in pilot study. The artifact rejection criterion was 80 uV. Rejected trials were transparently (to the subject) replaced, although the subject of course experienced the rejected trials. The EEG electrodes were referentially recorded but the EOG electrodes were differentially amplified. The forehead was grounded. Signals were passed through Grass P511K amplifiers with a 30 Hz low pass filter setting, and with high pass filters set (3 dB) at .1 Hz. Amplifier output was passed to a 12-bit Keithly Metrabyte A/D converter sampling at 125 Hz. For all analyses and displays, single sweeps and averages were digitally filtered off-line to remove higher frequencies; 3 dB point = 4.23 Hz. P300 was measured using the base to peak method (BASE–PEAK): The algorithm searches within a window from 400 to 900 ms for the maximally positive segment average of 104 ms. The pre-stimulus 104 ms average is also obtained and subtracted from the maximum positivity to define the BASE–PEAK measure. The midpoint of the maximum positivity segment defines P300 latency.

2.2.1. Task procedures

Each subject was run through nine blocks of at least 200 trials over three sessions within a 2 week period. Each of three different response style blocks was combined with three oddball stimulus types, congruent, neutral, and incongruent. By oddball, we refer to stimuli which are rare, especially meaningful, and requiring a unique response. The response styles were VERBAL—the subject spoke the color of the stimulus word out loud, BUTTON—the subject pressed one of three pre-memorized buttons corresponding to the three colors used in which stimulus words were displayed, and SILENT—the subject spoke the colors of the words silently to himself (herself). Behavioral reaction times (RTs) could be obtained only in the VERBAL and BUTTON blocks. In the former, a voice-actuated relay provided the response signal to the computer. ERPs were recorded during all blocks (even though it will be clear in the results that the “P300” obtained during the VERBAL block was largely artifactual). In each block, there was an oddball Stroop stimulus with $p=.15$ (30 trials in 200 trials). This was either 1) a color word in the congruent color, either blue, green, or gold, or, 2) in the incongruent blocks, one of those three color words but displayed in one of the other colors, or 3) a neutral non-color word (bread, pen, and town) displayed in one of the three same colors. Although we could have randomized to some extent the color in which neutral and incongruent stimuli were presented, we decided that since the congruent words were always presented in the same colors, we would also present incongruent and neutral words always in the same color. One frequent stimulus, the triad XXX, with $p=.85$, was randomly presented in one of the three colors used for other stimuli. On all trials, the task was to name the color either aloud, silently, or via a button press, as the response block required. Order of block experiences was counterbalanced except that each subject would always have one response style per 3-block day. The order of these days was counterbalanced across subjects. Within a subject, the order of stimulus type blocks was counterbalanced across the three days. Since all artifact rejected trials were replaced, there were at least 30 trials of averaged EEG activity for Stroop stimuli, and at least 170 trials for frequents.

For the SILENT blocks, the subjects were interrupted about every 30–40 trials, the program halted, and the subject asked to repeat the stimulus aloud. They were forewarned about these pop quizzes. Subjects failing to achieve 100% correct were to be dropped (there were none). For the other response styles, Subjects achieving <95% correct responses were dropped. There was only one of these.

2.3. Results

Fig. 1A and B show grand average P300s to all stimulus types from the Fz and Pz sites. In the SILENT and BUTTON conditions, there are distinct P300s with the Pz response > the Fz response, as is typical. (The Cz response, not shown, was intermediate.) The BUTTON P300 appears larger than in the SILENT condition, but there is a distinct P300 in the latter condition, which attests to the fact that the subjects were performing the task and noting the rare stimuli. Line graphs in Fig. 2A, B, and C, are consistent with these observations. Fig. 1A and B also indicate that in the non-VERBAL conditions, there does not appear to be a difference in latency for the Pz P300s. In contrast to results in the non-VERBAL conditions, Fig. 1C indicates that the “P300” at Fz is much larger than the response at Pz (note the large differences in amplitude legend keys), and that there indeed appears to be a reflection of the Stroop effect in the P300 latencies, especially at Fz, where the latency is earliest and the amplitude greatest for congruent stimuli, with the slowest latency and smallest amplitude for incongruent stimuli, with intermediate P300 attributes for neutral stimuli. These trends are also reflected at the other electrodes, though less obviously. Fig. 2D reflects these amplitude observations, indicating that the P300 scalp distributions for VERBAL responses are opposite to those for other response types for all three stimuli, and the

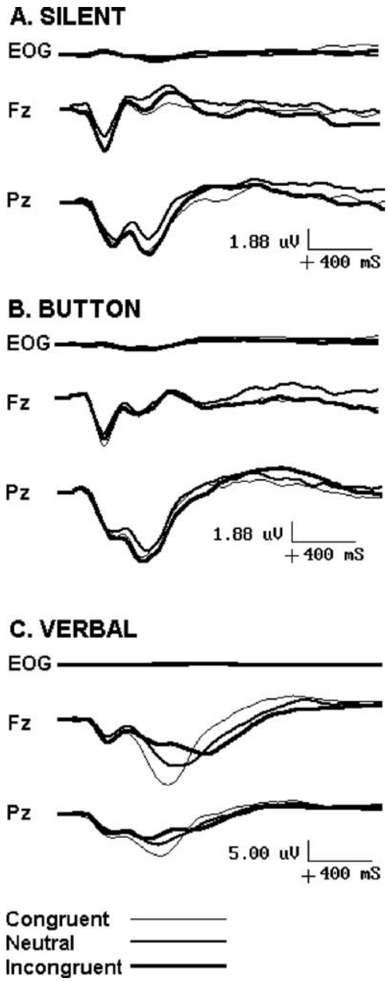

Fig. 1. Grand average ERP data from Fz and Pz to all stimuli and from all response conditions in Experiment 1 as indicated. Note difference in amplitude marker keys between VERBAL and other P300s. Positive down as shown for all ERP figures. These and all other ERP figures include the 104 ms pre-stimulus baseline as noted in the Methods section.

amplitudes are grossly larger—note the difference in amplitude range between the VERBAL and other P300s.

In support of these impressions, a 2 (response condition)-by-3 (site)-by 3 (stimulus type) ANOVA was performed on the P300 amplitude data of Fig. 2A, B and C. The response effect was significant; $F(1,11)=13.4$, $p<.005$ (BUTTON > SILENT), as was the site effect; $F(2,22)=8.23$, $p(\mathrm{GG})<.02$ (Pz>Cz>Fz.). ("GG" refers to the Greenhouse–Geisser probability, corrected for sphericity in repeated measure tests with $df>1$.) The stimulus type effect was not significant ($p>.1$), nor were any interactions (all $p>.4$). These results confirm the visual impression of the classical P300 scalp amplitude distribution of Pz>Cz>Fz having been obtained for all stimuli in the non-VERBAL conditions.

For the amplitude data of the VERBAL block (Fig. 2D, intentionally kept separate from the non-verbal blocks in which artifact-free data could be reasonably expected), we did a 3 (stimulus type)-by-3 (site) ANOVA. The stimulus type effect was marginally significant; $F(2,22)=3.6$, $p(\mathrm{GG})=.061$. Fig. 2D suggests that the incongruent vs. congruent data, especially anteriorly, carries this effect. The site effect is also significant, but as Fig. 2D makes clear, it is in the atypical direction: Fz>Cz>Pz; $F(2,22)=5.18$, $p(\mathrm{GG})<.05$. The interaction approached significance, $F(4,44)=2.7$, $p(\mathrm{GG})<.082$. A post-hoc 1×3 ANOVA on Pz only data failed to reach significance, $p>.1$. As noted above, qualitatively, the VERBAL "P300s" seemed abnormally large. Thus we compared the BUTTON (larger of the two non-VERBAL conditions) amplitudes with the VERBAL amplitudes at Pz only (where "P300s" were smallest in the VERBAL condition) for all stimuli combined in a 2 (response condition)-by-3 (stimulus type) ANOVA. Note that this comparison was conservative; the largest BUTTON value was compared with the smallest VERBAL value. Nevertheless the response condition effect yielded $F(1,11)=8.55$, $p<.02$. The stimulus type had no significant effect, $p(\mathrm{GG})=.27$. The interaction was not significant $p(\mathrm{GG})>.05$. In the interest of comparison with Ilan and Polich (1999) who used only a BUTTON block, we also examined the effect of stimulus type across all sites in our BUTTON condition: The effect on P300 amplitude was not in evidence, $F(2,22)=.395$, $p(\mathrm{GG})>.6$.

Fig. 3 contains line graphs of temporal data on RT (3A, VERBAL and BUTTON data), SILENT and BUTTON P300 latency at Pz (3B, from which non-distributional P300 attributes should be principally derived), and "P300" latency (3C) from all sites from the VERBAL block only. Obviously, there are no RT data from the SILENT condition, and since we expected and appear to have obtained a contaminated ERP response in the VERBAL condition (which is why we use quotation marks for the VERBAL "P300"), we show true Pz P300 data in Fig. 3B only in the non-VERBAL conditions, reserving Fig. 3C for the VERBAL latencies. The results are rather straightforward: Fig. 3A shows the classic Stroop effect with VERBAL responses, with the RTs in the order incongruent>neutral>congruent. This is also seen in the BUTTON condition, but the degree of Stroop facilitation (i.e., neural-incongruent) is, as typically reported, reduced. Fig. 3B does not suggest a reflection of Stroop effects in the true SILENT P300 latency at Pz. Finally, Fig. 3C does suggest the reflection of a Stroop effect in the "P300" latencies at the anterior sites, nearer the oropharyngeal structures, in the VERBAL conditions.

A 3 (stimulus type)-way ANOVA on the VERBAL RTs of Fig. 3A yielded $F(2,22)=56.3$, $p(\mathrm{GG})<.001$. To separately assess facilitation and interference, we also found that in the VERBAL condition, RT for incongruent stimuli>RT for neutral stimuli, $t(11)=6.7$, $p<.001$. Also, RT for neutral stimuli>RT for congruent stimuli, $t(11)=4.63$, $p<.002$. (Both these p-values satisfy a Bonferroni corrected post-hoc test criterion.) Thus both Stroop facilitation as well as interference were achieved. In the BUTTON condition, a 3-way ANOVA also yielded a significant main effect of stimulus type, $F(2,22)=22.3$, $p(\mathrm{GG})<.001$, however, only

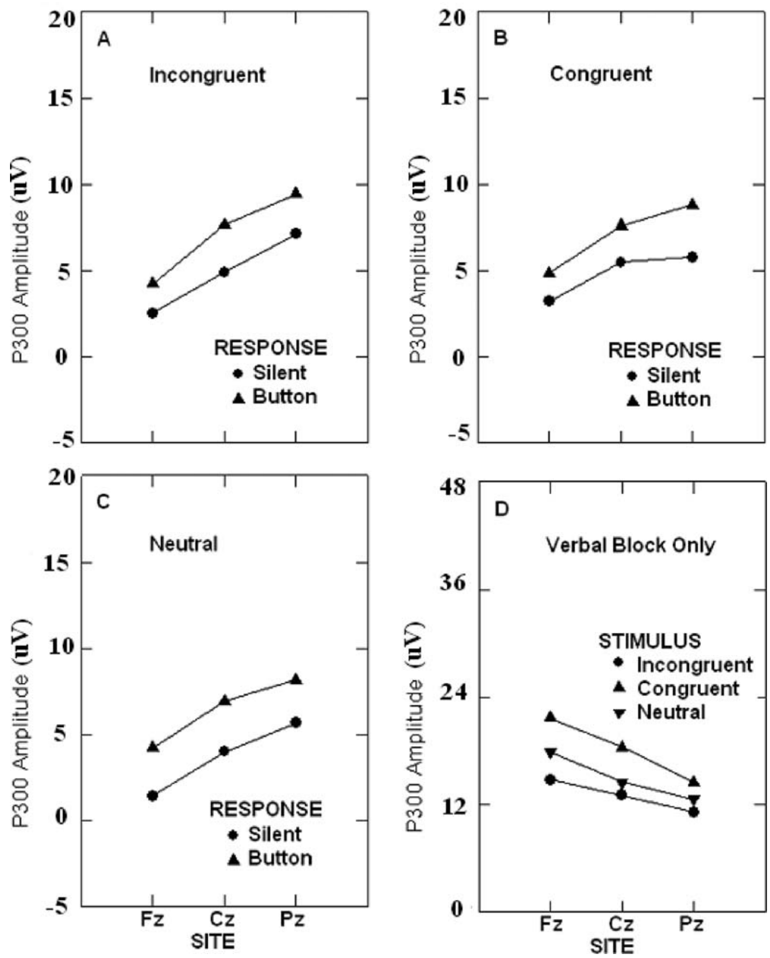

Fig. 2. Calculated P300 amplitudes for all sites and stimulus types in all response conditions in Experiment 1 as indicated. Amplitudes are in uV. Our SYSTAT 10 (SPSS) software graphs data means based on a file of individual values on which it calculates means and variances. The *y*-axes are automatically designed to illustrate the full range of individual values. Thus there may appear to be wasted space, but the space conveys the range used by individuals.

interference was demonstrated statistically, $t(11)=9.01$, $p<.001$. While the congruent RTs (628 ms)<neutral RTs (648 ms), this difference was not significant, $p>.5$.

Fig. 3B suggests no effect of stimulus type on P300 latencies at Pz, and this was confirmed by failures of 3-way ANOVAs to find main stimulus effects, both $p>.2$.

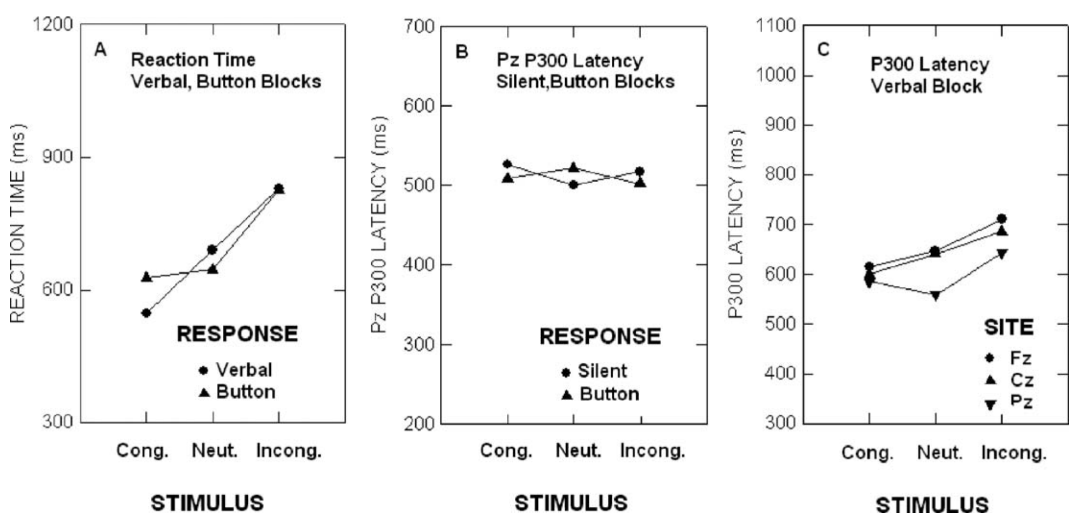

Fig. 3. RTs (A), P300 latencies (B, C, in ms) for all stimulus types in Experiment 1 as indicated. Pz P300 latencies are in B. A and B contain data from non-VERBAL response conditions; C contains VERBAL response data from all sites.

In the VERBAL block (Fig. 3C), there appears to be a reflection of a Stroop effect on the "P300" latencies at Fz and Cz. However, a 3(stimulus type)-by-3 (site) ANOVA yielded only a main effect of site, $F(2,22)=11.2$, p(GG)<.005. The stimulus type effect was $p>.3$, and the interaction was ns at $p>.15$. A post-hoc t-test on incongruent vs. congruent latencies at Fz yielded $t(11)=2.27$, $p<.05$. Similar results obtained at Cz: incongruent and congruent latencies differed, with $t(11)=2.26$, $p<.05$. At Pz, there were no significant latency differences. These findings of RT Stroop effects, but no Pz latency effects with verbal responding replicate Duncan-Johnson and Kopell (1981).

2.3.1. Discussion, Experiment 1: introduction to Experiments 2 and 3

This study utilized rare Stroop stimuli, and satisfied the P300 scalp distribution criterion of Pz>Cz>Fz in the BUTTON and SILENT conditions (Fabiani et al., 1987). Stroop interference and facilitation effects were seen in the RTs of VERBAL and BUTTON conditions, although facilitation, which is often reported to be the weaker of the Stroop effects (MacLeod, 1991), did not reach significance in the BUTTON blocks. Ilan and Polich (1999), in a button-press Stroop paradigm, did get significant facilitation effects using more than twice the number of subjects as we did. In our study, as in that of Ilan and Polich (1999), P300 latencies at Pz in non-VERBAL conditions showed no reflections of RT effects. These results continue to support the notion that Stroop effects are on the response selection, rather than the stimulus evaluation side of cognitive processing. However, the present results allow greater confidence than that found in previously published data that the latency indices used to support this inference were true P300 latencies, since a key paradigmatic antecedent of P300 elicitation, the low stimulus probability of Stroop stimuli, as reviewed in the introduction, were utilized here for the first time. Ilan and Polich (1999) did demonstrate the Pz>Cz>Fz distribution, supporting the likelihood that they were indeed recording true P300 data, even though all their stimuli were of equal probability (.25). A possible confound in their study was the fact that although their four congruent stimuli could appear in only four colors, their incongruent and neutral words appeared in 8 and 12 colors respectively. This difference between their and our studies might account for the fact that although they found an effect of stimulus type on Pz P300 amplitude, we did not in our BUTTON condition. Consistent with our results, they found no Pz latency correlates of RT Stroop effects.

The initial seminal paper utilizing P300 latency to analyze the Stroop effect (Duncan-Johnson and Kopell, 1981) was based on EEG data collected during vocalized responding (as in our VERBAL blocks). Neither were scalp distribution data presented; only Pz results were shown and mentioned. We believe Fz data from that study would have shown the same artifact-containing waveforms as those we presented during VERBAL blocks. Szirtes and Vaughan (1977) would have predicted the same artifact, which has also been considered by others (Ilan and Polich, 1999; West and Alain, 1999; Liotti et al., 2000; Atkinson et al., 2003) since our initial discussion of the artifact problem in Stroop studies in which SILENT blocks were first utilized (Rosenfeld and Woodley, 1994). Although the artifact problem appeared here to decrease from Fz to Pz (Fig. 2D), i.e., as the recording is taken from increasingly further away from the oropharyngeal structures (the source of the artifact, Szirtes and Vaughan, 1977), nevertheless the largest Pz response in our non-VERBAL blocks was 3.5 uV (significantly) less than Pz responses in our other response conditions, suggesting that VERBAL Pz data also contain moderate speech related artifact. The Fz and Cz latencies of "P300" indeed suggested a Stroop effect (Fig. 3C), which would suggest stimulus evaluation effects if one believed that these anterior waveforms to be free of artifact. In view of the preceding discussion, this strikes us as a mistaken conclusion.

Although Experiment 1 satisfied the low probability antecedent of P300 elicitation by Stroop stimuli, it did not contain explicit Stroop target stimuli, which indeed, were also not used by any previous ERP study of the Stroop effect. Experiment 2, which utilized both VERBAL and SILENT conditions, was designed to remedy this missing element. It was run on subject pool participants, perhaps not as cooperative as the advanced students run in Experiment 1. This could raise suspicions about participant cooperation in the SILENT blocks. Thus Experiment 3 was an attempted replication of the SILENT blocks in Experiment 2, except that advanced elective course students comprised the subject pool for the last study. We also did not use congruent stimuli in these last 2 studies, so that we could not study weaker facilitation, effects, but only Stroop interference of incongruent words relative to neutrals.

2.3.2. Methods, Experiments 2 and 3

These studies are treated together since the latter was a partial replication of the former.

2.3.3. Subjects

For Experiment 2, 20 students from the introductory psychology pool were recruited into two groups, neutral and incongruent, on the basis of which kind of stimuli they viewed. Data from one subject in the latter group were lost. For Experiment 3, a completely repeated measures design, 8 subjects from an advanced, elective psychophysiology lab course were run along with four friends/associates; $n=12$. All subjects had normal or corrected to normal vision (no contact lenses), and signed an IRB-approved consent form.

2.3.4. Recording procedures

These were identical to those in Experiment 1.

2.3.5. Task procedures

All subjects in incongruent blocks saw the word Blue, randomly presented in either Green or Purple colors, the word Purple, randomly presented either in Blue or Green colors, and the word Green, randomly presented in either Purple or Blue colors; the latter was designated the target with $p = 1/6$. In neutral blocks, subjects saw the words Book, randomly in either Green or Purple, Pencil in either Green or Blue, and Glove in either Purple or Blue; the latter was designated target also with $p = 1/6$. Color experiences were thus all the same in neutral and incongruent blocks. The task on each trial was to first name the color, followed by signaling of target ("Yes") or non-target ("No"). We chose this response order to prevent the target discrimination task from interfering with the color-naming task. There were two types of response blocks used in Experiment 2, SILENT and VERBAL. In Experiment 2, neutral and incongruent blocks were given to two separate groups, with $n = 9$ and 10, respectively. In each group the SILENT and VERBAL block orders were counterbalanced across subjects. In Experiment 3, only the SILENT condition was run but all subjects received both incongruent and neutral stimuli, with counterbalanced orders across all 12 subjects.

2.3.6. Results

Fig. 4 shows the grand averages from Experiment 2 in SILENT and VERBAL conditions for both incongruent and neutral stimuli, target superimposed on non-target, for all sites. In the SILENT conditions there appear to be normal P300 responses with target P300s > non-target P300s, and the classical P300 scalp distribution (Pz > Cz > Fz). In the VERBAL conditions, these results are not seen, and there appear to be no meaningful or consistent differences between target and non-target responses, and no site effects. Fig. 5, a set of line graphs based on data from Fig. 4, shows these results more dramatically. P300 amplitudes appear not to differ except in the SILENT/neutral blocks.

Regarding Fig. 5A, a 2(target–nontarget)-by-3 (site) ANOVA was done on the SILENT/incongruent data, Experiment 2: The target effect yielded $F(1,8) = 11.2$, $p < .02$. The site effect yielded $F(2,16) = 36.3$, $p(\mathrm{GG}) < .001$. Reflecting the steeper target than non-target curve, the interaction yielded $F(2,16) = 11.1$, $p(\mathrm{GG}) < .006$. This confirmed the visual impression of classic target and distribution effects. An analogous ANOVA on the SILENT/neutral P300s (Fig. 5B) yielded similar results: $F(1,9) = 31.9$, $p < .001$ for the target effect, $F(2,18) = 8.36$, $p(\mathrm{GG}) < .02$ for the site effect, and $F(2,18) = 7.69$, $p(\mathrm{GG}) < .02$ for the interaction. In contrast, for the VERBAL/incongruent data (Fig. 3C), there were no main effects ($p > .1$), but the interaction was $F(2,16) = 5.2$, $p(\mathrm{GG}) < .05$. For these doubtlessly artifact-influenced VERBAL data, it seems pointless to make additional interpretations. For the VERBAL/neutral data, again, there were no main effects, but a significant interaction, $F(2,18) = 9.01$, $p(\mathrm{GG}) < .02$. As for Experiment 1, one may conclude that in Experiment 2, the SILENT but not VERBAL P300s showed classic target and distribution effects.

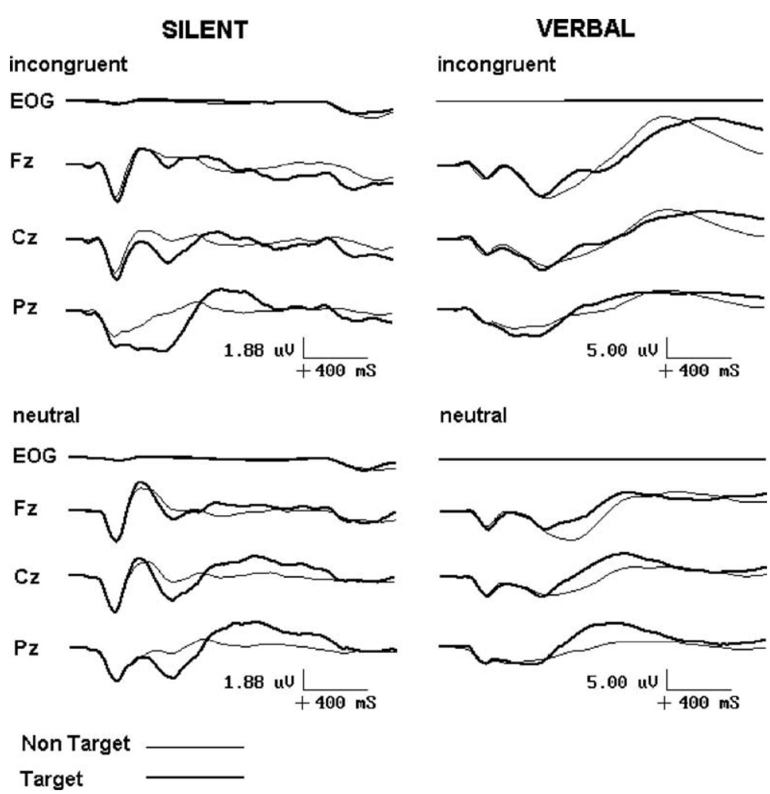

Fig. 4. Grand averages from Experiment 2 as in Fig. 1.

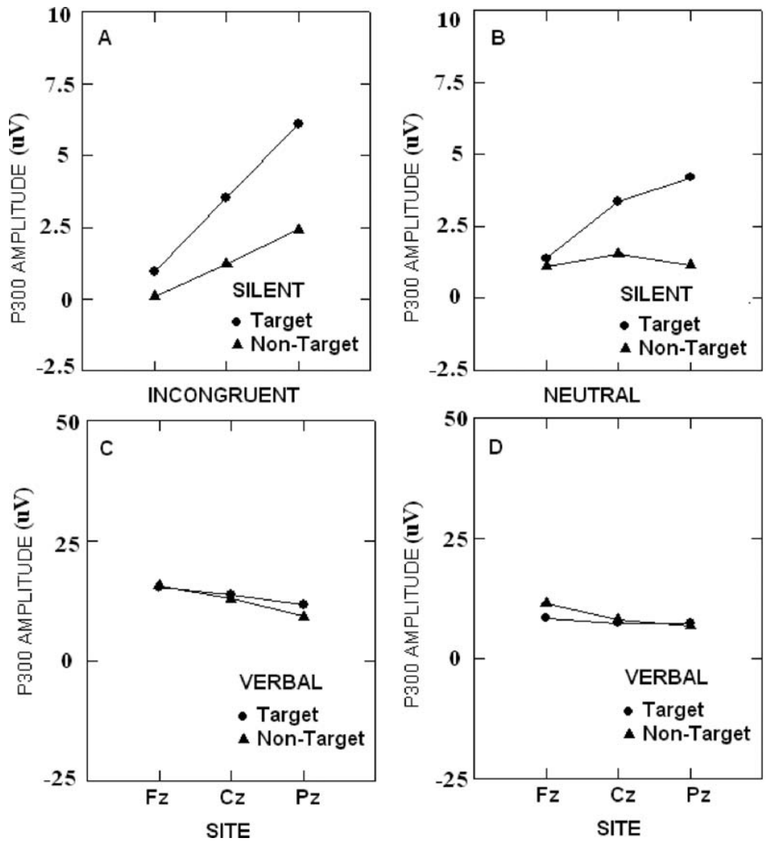

Fig. 5. Experiment 2 computed P300 amplitude data from all sites in SILENT response condition (A, B) and VERBAL response condition (C, D) in response to target and non-target stimuli. The left column figures (A, C) are for incongruent stimuli, the right column figures (B, D) are for neutral stimuli. Units: uV.

Fig. 6A and B show corresponding RTs and Pz P300 SILENT latencies. As in Experiment 1, there is a clear Stroop interference effect seen with RTs (6A, both for targets and non-targets), but not with Pz latencies in the SILENT condition (6B). Fig. 6C from Experiment 3 replicates the lack of latency effects seen in Fig. 6B from Experiment 2. To save space, grand averages from Experiment 3 are not shown, but the amplitude results were similar to what is seen in Fig. 4 (Experiment 2), and the line graphs of Fig. 7A and B from Experiment 3 show classical target and distribution effects.

In the all SILENT Experiment 3 (Fig. 7), a 2(target/non-target) $\times$ 3(site) ANOVA (completely within-subject) yielded a significant target effect, $F(1,11)=10.47$, $p<.009$, as well as a significant site effect, $F(2,22)=16.9$, $p(\mathrm{GG})<.001$, but no interaction ($p>.8$) for the incongruent data(Fig. 7A). An analogous ANOVA on neutral data (Fig. 7B) yielded parallel results, with $F(1,11)=13.1$, $p<.004$ for target effect, $F(2,22)=9.08$, $p(\mathrm{GG})<.009$ for the site effect and no interaction $p>.37$. Again, these are the classic P300 target and distribution effects. For the sake of completeness, although we had no hypothesis about P300 latency as a function of site and target condition, we did parallel 2×3 ANOVAs on the neutral and incongruent latency data (Fig. 7C and D). There were no significant effects, as suggested by Fig. 7C and D.

With reference to Fig. 6A, a 2(neutral vs. incongruent group)-by-2 (target vs. non-target) ANOVA yielded classic Stroop interference in the VERBAL RT data of Experiment 2, $F(1,17)=17.47$, $p<.002$. The target effect was ns at $p>.8$, and the interaction ns at $p>.13$. For the SILENT Pz latency data (Fig. 6B), there were no significant effects (all $p>.18$) in an analogous ANOVA. These last results were replicated in Experiment 3 (Fig. 6C), with all $p>.5$. It is also noteworthy that in Fig. 6B, neutral latencies were (insignificantly) longer than incongruent latencies, the reverse of the classic Stroop interference seen in the significant RT results.

2.3.7. General discussion

These studies showed consistently that Stroop effects seen with RT are not reflected by Pz P300 latency data. Results in the three different response modalities were as expected: VERBAL: Clear Stroop effects on RT were seen but the effects on P300 latency could not be accurately assessed due to artifact in this modality. Scalp distribution data demonstrated the artifact. In the BUTTON modality, Stroop effects were found with RT, with no reflection in P300 latency, and the P300 amplitudes showed the classic P300 scalp distribution and target effects. In the SILENT modality, the results were as in the BUTTON modality except RT effects could not be obtained. Thus, for the first time, Stroop effects were seen in both target and non-target

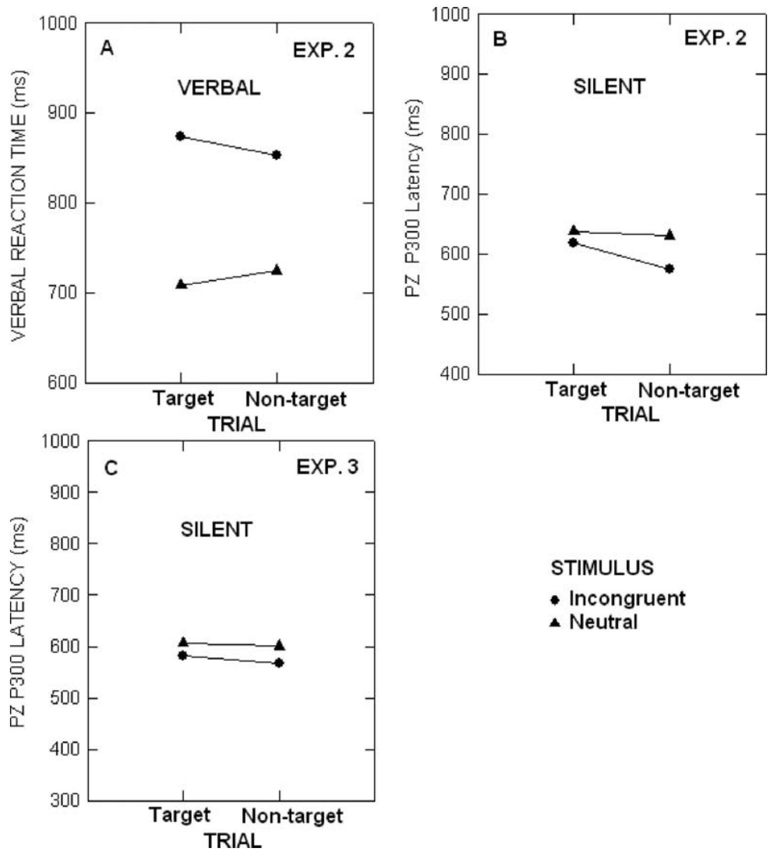

Fig. 6. Experiment 2 VERBAL RTs (A) and SILENT Pz P300 latencies (B) in ms in response to target and non-target stimuli. C is the same as B, except the data are from the Experiment 3 replication.

trials (Experiment 2), even as the Stroop interference was occurring (Fig. 6A), and also for the first time, the dissociation of RT and P300 latency was shown with Stroop stimuli having classically low probability in an oddball paradigm, that is, a paradigm with one type of rare, meaningful stimulus requiring a unique response. We built on the well-known finding of Duncan-Johnson (1981) that P300 latency may be taken as a metric of stimulus evaluation time, and utilized the seminal conceptualization of Duncan-Johnson and Kopell (1981) that correlation of RT and P300 latency could shed light on the locus of Stroop effects in cognitive processing. Consistent with their original study, we found consistent support for the response selection hypothesis. This support is also consistent with results of other studies sensitive to the issues of speech artifact and scalp distribution which Duncan-Johnson and Kopell (1981) did not directly address (e.g., Ilan and Polich, 1999; West and Alain, 1999; Liotti et al., 2000; Atkinson et al., 2003). While we are confident that our present protocols were the most certain to have elicited classical P300 ERPs, it remains likely that other recent studies (e.g., Ilan and Polich, 1999) demonstrating the classical P300 scalp distribution also provided strong support for the response selection hypothesis of the Stroop effect. Indeed, neither an explicit target nor a low objective probability event are absolutely necessary to elicit P300 (Duncan-Johnson and Donchin, 1977; Johnson, 1993). It is likely, however, that these attributes increase stimulus salience or meaningfulness, which appear to maximally involve the cognitive mechanism whose activity is signaled by P300. The presently reported data thus provide the strongest support for the response selection hypothesis.

It might be objected that possibly, only with button-press responses can one be certain of the relationship of RT and P300 latency. This would follow from the fact that 1) clean P300s cannot be recorded, given the present state of the art, during speech, and 2) a silent responding option makes it ultimately impossible to know whether or not the subject is performing the task as instructed. On the other hand, since the verbal RT typically produces much more robust Stroop effects than manual responses (MacLeod, 1991), it has been argued that verbalized RTs are more likely than manual RTs to be tapping into the true sources of Stroop phenomena. Moreover, we would argue that beyond scalp distribution, our demonstrable P300s and target effects during silent responding blocks provide strong confirmation, not previously available, that subjects were indeed performing the target discrimination tasks silently as instructed. We lack equivalently powerful evidence that simultaneous color naming was occurring, but it appears reasonable to assume that if a subject is performing one task silently (the oddball task), he is

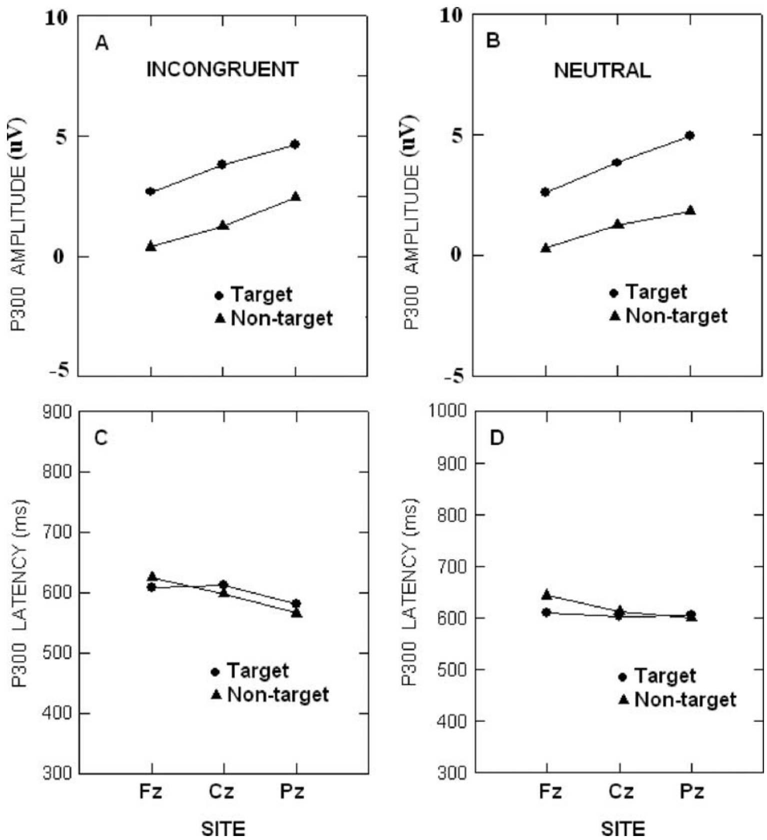

Fig. 7. Computed target and non-target amplitude (A, B; uV) and latency (C, D; ms) data from all sites in Experiment 3. A, C are for incongruent stimuli; B, D for neutral stimuli.

probably also cooperating with the other (color naming) task. This view is also buttressed by the fact that advanced psychology majors in an elective laboratory course were utilized as subjects here in the first and third studies. We do suspect that not all subjects were equivalently involved in the silent blocks. We base this view on 1) the significantly larger P300s seen in BUTTON blocks than in SILENT blocks in Experiment 1 for all stimulus types, at all sites, and 2) our informal observations, not previously noted, of small or absent P300s for the SILENT blocks of a small minority of subjects in all three studies. Perhaps this minority of subjects was unable to maintain alertness without awareness of task enforcement. Nevertheless, this minority clearly did not preclude our observation of the major findings presented here. These findings allow the conclusion that with manipulations designed, for the first time, specifically to elicit the P300 component of the ERP, namely, using target Stroop stimuli of low probability, P300 latency as a function of stimulus type-congruent, neutral, and incongruent, unlike RT, is flat. This implicates the locus of Stroop interference in response selection mechanisms. One caveat about this conclusion would concern our lack of use of higher density electrode arrays such as those used in recent studies by West et al. (2004) and West (2003) whose resolution might have allowed removal of possible effects of other components overlapping P300, with differing results.

References

Atkinson, C.M., Drysdale, K.A., Fulham, W.R., 2003. Event-related potentials to Stroop and reverse Stroop stimuli. International Journal of Psychophysiology 47, 1–21.

Donchin, E., Coles, M.G., 1988. Is the P300 component a manifestation of context updating? Behavioral and Brain Sciences 11, 357–374.

Donchin, E., Ritter, W., McCallum, C., 1978. Cognitive psychophysiology: the endogenous components of the ERP. In: Callaway, E., Tueting, P., Koslow, S. (Eds.), Brain Event-Related Potentials in Man. Academic Press, New York, pp. 349–441.

Duncan-Johnson, C.C., 1981. P300 latency: a new metric of information processing. Psychophysiology 18, 207–215.

Duncan-Johnson,, Donchin,, 1977. On quantifying surprise: the variation of event related potentials with subjective probability. Psychophysiology 14, 456–467.

Duncan-Johnson, C.C., Kopell, B.S., 1981. The Stroop effect: brain potentials localize the source of interference. Science 214, 938–940.

Fabiani, M., Gratton, G., Karis, D., Donchin, E., 1987. The definition, identification, and reliability of measurement of the P300 component of the event-related potential. In: Ackles, P.K., Jennings, J.R., Coles, M.G.H. (Eds.), Advances in Psychophysiology, vol. 2. JAI Press, Inc., Greenwich, CT.

Ilan, A.B., Polich, J., 1999. P300 and response time from a manual Stroop task. Clinical Neurophysiology 110, 367–373.

Johnson, R., 1993. On the neural generators of the P300 component of the event-related potential. Psychophysiology 30, 90–97.

Kerns, J.G., Cohen, J.D., MacDonald, A.W., Cho, RY., Stenger, V.A., Carter, C.S., 2004. Anterior cingulate conflict monitoring and adjustments in control. Science 303, 1023–1026.

Kutas, M., McCarthy, G., Donchin, E., 1977. Augmenting mental chronometry; the P300 as a measure of stimulus evaluation time. Science 197, 792–795.

Liotti, M., Woldorff, M.G., Perez III, R., Mayberg, H.S., 2000. An ERP study of the temporal course of the Stroop color–word interference effect. Neuropsychologia 38, 701–711.

MacDonald, A.W., Cohen, J.D., Stenger, V.A., Carter, C.S., 2000. Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. Science 288, 1835–1838.

MacLeod, C.M., 1991. Half a century of research on the Stroop effect: an integrative review. Psychological Bulletin 109 (2), 163–203.

Rosenfeld, J.P., Woodley, S.J., 1994. True oddball P300 latency does correlate with reaction time in a Stroop task. Psychophysiology 31, S83 (Supplement, abstract).

Spencer, K.M., Dien, J., Donchin, E., 2001. Spatiotemporal analysis of the late ERP responses to deviant stimuli. Psychophysiology 38, 343–358.

Stroop, J.R., 1935. Studies of interference in serial verbal reactions. Journal of Experimental Psychology 18, 643–662.

Szirtes, J., Vaughan Jr., H.G., 1977. Characteristics of cranial and facial potentials associated with speech production. Electroencephalography and Clinical Neurophysiology 43 (3), 386–396.

West, R., 2003. Neural correlates of cognitive control and conflict detection in the Stroop and digit-location tasks. Neuropsychologia 41, 1122–1135.

West, R., Alain, C., 1999. Event-related neural activity associated with the Stroop task. Cognitive Brain Research 8, 157–164.

West, R., Bowry, R., McConville, C., 2004. Sensitivity of medial frontal cortex to response and nonresponse conflict. Psychophysiology 41, 739–748.