Different time courses of Stroop and Garner effects in perception — An Event-Related Potentials Study

Lars T. Boenke a,b, Frank W. Ohi b,c, Andrey R. Nikolaev a,d, Thomas Lachmann a,e, Cees van Leeuwen a,*

a Laboratory for Perceptual Dynamics, BSI RIKEN, 2-1 Hirosawa, Wako-shi, 351-0198 Japan
b Leibniz Institute for Neurobiology, Magdeburg, Germany
c Otto-von-Guericke University, Magdeburg, Germany
d Institute of Higher Nervous Activity, Moscow, Russia
e University of Kaiserslautern, Kaiserslautern, Germany

A B S T R A C T

Visual integration between target and irrelevant features leads to effects of irrelevant feature congruency (Stroop) or variation (Garner) on target classification performance. Presenting closed geometrical shapes as stimuli, we obtained Stroop and Garner effects of one part of their contour on another, in response times and error rates. The correlates of these effects in brain activity were observed in event-related potentials (ERP). Stroop effects occurred in ERP amplitude of the N1 and N2 components, starting about 170 ms after stimulus onset; Garner effects occurred in amplitude of the rising part of the P3 component, starting about 330 ms after stimulus onset. A subsequent point-wise analysis of Stroop and Garner effects in ERP showed that they belong to different, cascaded processing stages. The difference in time course between Stroop and Garner effects in ERP is in accordance with the view that both are produced by different mechanisms, the former sensitive to interference within presentations and the latter sensitive to interference between presentations. The brief interval of 330–370 ms after stimulus onset when these two mechanisms overlap may correspond to the central processing bottleneck, responsible for the combinations of Stroop and Garner effects generally found in response times.

© 2009 Elsevier Inc. All rights reserved.

When the visual system receives information, separate features such as line and contour segments are initially being detected, only to rapidly become integrated into the structures, textures, shapes, and objects that enter our awareness. During this process, attentional selection mechanisms operate both on the level of features and whole objects. The function of these mechanisms is to prevent intrusion into awareness of irrelevant information (Desimone and Duncan, 1995; Mounts and Tomaselli, 2005). Selective attention cannot, however, prevent the intrusion of information that visual feature integration has already tied up with the target.

Intrusion of irrelevant information can therefore be read as a sign that feature integration has taken place (cf. Mounts and Tomaselli, 2005; Pomerantz and Lockhead, 1991). Such intrusions can be observed in behavior, in particular as they lead to Stroop and Garner effects (Stroop, 1935; Garner 1974, 1976, 1988). These effects have therefore played a vital role in behavioral studies of feature integration (Pomerantz et al., 1989; MacLeod, 1991; van Leeuwen and Bakker, 1995; Patching and Quinlan, 2002). Our present study will contrast these effects, using the high temporal resolution of the human scalp EEG to measure their time course.

In the classical Stroop task, naming the ink color of a color-word is delayed if the color-word is different (incongruent) from the color of the ink which has to be named, e.g., the word red printed in green ink (Stroop, 1935). Stroop effects reported in the literature vary widely (MacLeod, 1991); this has led to several naming controversies. According to the most encompassing definition (Irtel, 1993), the Stroop effect is a deterioration of a response to a target feature resulting from one or more incongruent but irrelevant other features presented at the same trial, as compared to congruent features. Thus, Stroop effects extend to the auditory domain (Hamers and Lambert, 1972), the Eriksen flanker paradigm (Eriksen and Eriksen, 1974), occur between individual names and faces (Egner and Hirsch, 2005), numerical values and physical sizes (Algom et al., 1996), names of countries and their capitals (Dishon-Berkovits and Algom, 2000), and object- or shape-based stimuli (Pomerantz et al., 1989; van Leeuwen and Bakker, 1995). In the latter, the Stroop effect is often also referred to as the congruency effect (for a review: Marks, 2004; Patching and Quinlan, 2002).

The congruency variant of the Stroop task is the focus of our study. Consider the four stimuli from Fig. 1. They are characterized by a combination of inner and outer contours that qualify two of them as congruent and the other two as incongruent. For each figure, participants responded whether its inner contour has a rectangular or triangular shape, irrespectively of the outer contour. A resulting
congruency effect implies, according to our diagnostic, that inner and outer contours were integrated into a perceptual whole.

Garner interference was named by Pomerantz (1983) after Garner (1974, 1976, 1988). According to Garner, the dimensions of an abstract “feature space” are called “integral” if irrelevant feature variation along one dimension affects the perception of the other, as with, for instance, brightness and saturation (Garner, 1976); otherwise they are considered separable, as circle size and radius inclination (Garner and Felfoldy, 1970). Experimentally observing Garner interference requires a design comparing blocks of trials with the irrelevant feature held constant (baseline condition), varied randomly (filtering condition), or covaried with the target dimension (correlation condition). If performance in the filtering condition differs from baseline and/or correlated conditions, this indicates integral dimensions (Melara and Algom, 2003).

Garner (1988) applied this method to graphemic stimuli with the dimensions “letters” and “color”. Letters C and O were presented in green or red ink color. The task was to name the ink color, which varied randomly in both letter conditions. Here, the irrelevant feature was associated with the “letters” dimension. In the baseline condition, the letters “O” or “C” would occur in separate blocks; in the filtering conditions they would be randomly intermixed. Irrelevant variation of the letters slowed down the response to the color dimension, which implies that letter identity and color are integral dimensions.

In this study we observed the Garner effect of irrelevant variation of the outer contour while responding to the inner one in the stimuli of Fig. 1. Garner effects were taken as another indication, besides the Stroop effect, that both contours were integrated into a perceptual whole.

As both Stroop and Garner effects deal with perceptual integration, intuitively speaking, they might be expected to occur jointly. On the other hand, a perceptual whole is not created instantaneously; Stroop and Garner effects may occur at different moments during the process of feature integration. The timing issue has been addressed in behavioral studies using Sternberg’s (1969) additive factors logic. According to this logic, when two factors are orthogonally varied, their process distinctions but no corresponding prescription for eliminating instances of full dissociation have been observed (Melara and Mounts, 1993; van Leeuwen and Bakker, 1995; Patching and Quinlan, 2002). These studies suggest a difference in time frames: Stroop effects tend to be predominant when conflicting information is presented within a narrow time window: to produce interference congruent and incongruent features have to be presented within an approximately 100 ms time frame (Flowers, 1990). Thus, memory involvement is minimal. The Garner effect, on the other hand, involves memory, as conflicts occur between presentations.

The double dissociation method is not a fail-safe procedure for making process distinctions. According to Van Orden et al. (2001), this method is liable to false positives: it has a prescription for adding process distinctions but no corresponding prescription for eliminating ones. To reach more reliable conclusions, we returned to the issue of stages. We recruited a method with sufficient temporal resolution to distinguish cascaded stages. The present study used event-related potentials (ERP) to reveal the neural correlates of Stroop and Garner effects on brain activity during perceptual classification of the stimuli of Fig. 1. The difference in the way memory is supposedly involved between Stroop and Garner effects, led van Leeuwen and Bakker (1995) to predict that Stroop effects would arise earlier within the time course of perception than Garner effects. From a dynamic, interactive point of view, interference effects take more time to emerge the larger the feedback loop involved. Stroop effects, therefore, will arise faster than Garner, as memory processes are largely kept out of the loop. Thus, Stroop effects should be manifested earlier in ERP than Garner.

Efforts to compare Stroop and Garner in an ERP experiment were recently made in Caclin et al. (2008) and earlier in Lew et al. (1997). Caclin et al. (2008) evaluated Stroop-like congruency and Garner effects in the auditory domain, using timbre dimensions. Their time course was estimated in a point-wise analysis of ERP waves. The authors observed the first neural correlate of congruency about 55 ms after stimulus onset. A Garner effect occurred later, about 250 ms after stimulus onset. This result corresponds to the theoretical prediction that the Stroop effect has to precede the Garner effect (van Leeuwen and Bakker, 1995) and licenses similar predictions for the visual domain. In one condition, Caclin et al. (2008) obtained a very early Garner effect (at about 30 ms after stimulus). These authors used timbre dimensions that were difficult to discriminate. So it is possible that these results reflect investment of attentional resources. Our stimuli, by contrast, do not require excessive attentional efforts in the filtering condition.

Lew et al. (1997) used an auditory task, asking their participants to determine the gender of a male or female voice speaking the words...
“MUMMY” or “DADDY”, as well as a visual task, using the classical Stroop paradigm. They obtained mixed ERP results. In the auditory modality, Stroop conditions affected brain signals early, in N1 amplitude, and Garner conditions later, in P3 amplitude. In the visual modality, a Stroop effect was reflected only in the late waveform about 450–850 ms after stimulus onset but a Garner effect was found in the amplitudes of N140, P3, and in the 450–850 ms interval. The authors concluded that both Stroop and Garner effects occur in several perceptual and post-perceptual stages of processing.

Such a mosaic of findings may also be explained as a result of mixing semantic and perceptual (auditory or visual) dimensions. Our brain resolves conflicting information differently depending on context and attentional resource allocation (Cohen and Shoup, 2000). Caclin et al. (2008) proposed that neural mechanisms (see, for instance, a recent meta-analysis in Nee et al., 2007). We may distinguish tasks in which congruency occurs within a feature domain, i.e., the flanker task (Eriksen and Eriksen, 1974), Navon task (Navon, 1977), or object–shape tasks (Pomerantz, 1983; van Leeuwen and Bakker, 1995), versus those where congruency occurs between different feature domains (such as word meaning and ink-color, i.e., the classical Stroop task). In the first case target and non-target will compete for the same processing resources from an early stage; in the latter case competition will arise only after the target and non-target have been processed up to some point in different processing streams (Aine and Harter, 1984a,b; Treisman, 1991; Cohen and Shoup, 2000). Caclin et al. (2008) proposed that congruency across different domains may still show early effects in ERP, as long as these dimensions reflect basic, semantics-free, perceptual features. At least in visual processing, however, early semantic influences cannot be excluded for basic features (Hochstein and Ahissar, 2002).

Tasks in which congruency occurs between different domains are thus optimally suitable to reveal early correlates of Stroop effect in ERPs. The ink color from the visual and the color word from the semantic domain in classical Stroop stimuli, for instance, are likely to be processed through their distinct channels and at unequal rates. In Aine and Harter (1984a,b) color-related ERP modulation occurred consistently earlier than word related processing. Due to anatomical constraints, the physiological correlates of convergence between semantic and visual processing streams were detected only in late ERP components, often in frontal and central areas (Rebai et al., 1997; West and Alain 1998; Ilan and Polich 1999; Liotti et al., 2000). On the other hand, if the relevant and irrelevant features stem from the same domain they would more likely compete for same processing channel. Interference may arise early within such a channel, giving rise to an early ERP correlate of the Stroop effect. Therefore, we chose target and nontarget features from the same domain, in order to optimize the likelihood of observing early correlates of the Stroop effect.

Another restriction in the visual domain is imposed by spatial properties of attention: a non-target feature may be placed in the same (as in the classical Stroop task) or different location from the target (as in the flanker task). Stroop interference can be observed for naming the color of a geometrical target-shape flanked by an irrelevant color–word printed in grey (MacLeod, 1991; Morein-Zamir et al., 2002) depending on distance between target and flanker (Shalev and Algom, 2000). We therefore presented target and distracter foveally within the same narrow spatial range and on one object, in order to minimize spatial attention effects.

Spatial attention plays a modulatory role in the well-known Navon, or local/global, interference task (Navon, 1977). Consider a stimulus of the Navon type, e.g., a global “N” composed of local “Ns” (the congruent case) or of “W”s (the incongruent case). Whereas spatially distributed attention facilitates processing of the global level, focused attention facilitates processing of the local level. Whichever of the two levels is preferred, the interference from the non-preferred to the preferred level will be reduced. Global vs. local processing preference, therefore, acts as an intervening variable in the competition between target and non-target features. The preference is highly context-dependent (Collin et al., 2003). To exclude such intervening variables, we assured that target and non-target are part of the same whole, a closed geometrical shape.

In our stimuli the global non-target feature is somewhat larger and more prominent in shape than the local target feature. Melara and mounts (1993) demonstrated with the classical Stroop stimuli that larger salience of the irrelevant dimension was needed for the Stroop effect to occur. We chose to use difference in salience between target and non-target features to maximize the interference. Our choice was based on the earlier work in which van Leeuwen and Bakker (1995) studied different (global or local) roles of targets and non-targets using similar triangular or rectangular stimulus features. We used the feature combination which was found to be suitable in that study for emergence of both Stroop and Garner effects.

For the stimuli in Fig. 1, we expect slower responses for the incongruent than for the congruent condition (Stroop effect) and slower responses for the filtering condition than for the baseline condition (Garner effect). We expect ERP correlates of the Stroop effect to appear earlier than those of the Garner effect. To test this central hypothesis we used two complementary methods of analysis: the study of ERP components with well-established physiological meanings and pointwise tests to determine the onset latencies of Stroop and Garner effects.

For the classical ERP components, as our stimuli were designed to test perceptual integration, we expected a modulation of the early perceptual components P1 and N1. Such effects were previously observed in studies of interference in the auditory domain (Lew et al., 1987; Kaganovich et al., 2006; Caclin et al., 2008). P1 is an early perceptual component with peak latency about 80–130 ms which is manifested in the lateral occipital areas (Hillyard et al., 1998). P1 is also known as the earliest manifestation of top-down influence during sensory processing (Klimesch et al., 2007). N1 is a perceptual component with peak latency about 150–200 ms in parietal and lateral occipital areas (Hillyard et al., 1998; Luck, 2005), sensitive to spatial attention (reviewed in Mangun, 1995; Hillyard et al., 1998). It may indicate discrimination processes at the attended location (Mangun and Hillyard, 1991; Vogel and Luck, 2000). Our target and nontarget features, although spatially distinct, are located closely together. N1 effects of congruency are therefore likely to reflect discrimination processes at the attended location. We expect higher N1 amplitude in the incongruent than in the congruent condition assuming that incongruent target discrimination is more perceptually effortful than congruent target discrimination.

Next, we predict effects on the component N2. A recent review provided evidence that the N2 encompassed a family of ERP responses with similar latencies (about 200–350 ms) but different origins (Folstein and Van Petten, 2008). Two anterior N2 subcomponents can be distinguished which relate, respectively, to perceptual template mismatch detection and to cognitive control; a posterior subcomponent is related to visual attention in oddball tasks. N2 was shown to be sensitive to perceptual overlap between targets and non-targets (Nieuwenhuis et al., 2004; Folstein and Van Petten, 2004; Azizian et al., 2006). Perceptual overlap may occur between the visually similar target and non-target features of our stimuli. Perceptual overlap is most likely to invoke cognitive control associated with the fronto-central N2 subcomponent. We selected the fronto-central areas for evaluation of N2. We expect that the incongruent condition will result

2 The time window used for N2 widely varies in different studies, for example, 155–370ms (Heil et al., 2000), 200–400ms (Nieuwenhuis et al., 2004), 190–370ms (Folstein et al., 2008) depending on the method used.
in larger N2 amplitude than the congruent condition, as the inappropriate response has to be suppressed.

The last component of our interest is P3. This component with a peak latency about 400 ms in the visual domain may be divided into an anterior subcomponent P3a, related to the frontal attention mechanism, and a posterior subcomponent P3b in the parietal areas associated with attention and subsequent memory processing (reviewed in Kok, 2001; Polich, 2007). Studies of the Stroop combined with Garner effects (Lew et al., 1997; Caclin et al., 2008) and Stroop effect in isolation (Duncan-Johnson and Kopell, 1981; Ilan and Polich, 1999) revealed involvement of P3. As the Garner effect involves memory at the level between trials, we expected that the Garner effect may influence P3 more than the Stroop effect. Memory processes are reflected in a negative slow wave activity, which may lower the P3 amplitude (Wijers et al., 1989). We therefore assume that the P3 subcomponent relevant to our study is the memory-related P3b in the centro-parietal areas, which is expected to have more negative amplitude in the filtering condition than in the baseline condition. Our expected effects on N2 and P3 differ from the sustained negativity observed in the preceding Garner studies (Kaganovich et al., 2006; Caclin et al., 2008).

We expect Stroop and Garner to belong to different, but cascaded stages. There may thus be moments in time where each effect occurs alone and intervals where they overlap. Subsequent pointwise tests were performed to identify these stages.

Materials and methods

Participants

Nineteen healthy participants (13 male) aged between 17 and 33 years (mean age 25.8; SD: 4.5) took part in the experiment. All participants were right-handed and reported normal or corrected-to-normal vision. Participants signed a written consent form. For the participants were right-handed and reported normal or corrected-to-normal vision. Participants signed a written consent form. For the participants was achieved. We discarded the data from 3 out of 19 participants was achieved. We discarded the data from 3 out of 19 participants. The ethical committee of Otto-von-Guericke University in Magdeburg had given its permission for the study.

Stimuli

The stimuli were geometrical shapes with a closed contour which consisted of a spatially more extended outer part (called the global feature G) and a spatially more confined inner part (called the local feature L). Their size ratio was 2.5:1. The shape of both G and L contour could be either triangular or rectangular, resulting in the four possible combinations depicted in Fig. 1. In the congruent figures, both G and L were either triangular (G3L3) or both rectangular (G4L4). In incongruent figures G and L features differed (either G3L4 or G4L3). Each figure covered approximately 3.2° of the horizontal and 1.6° of the vertical visual angle. The figure contour was presented in black against a grey background (yielding a Michelson contrast of 94%).

Procedure

Participants were seated comfortably in a dark room at a distance of 1.7 m from the monitor. The stimuli were presented on a TFT monitor using “Presentation” software (Neurobehavioral Systems, Inc. Albany, CA).

A trial started with a blank screen followed by presentation of a fixation cross in the center of the screen. Blank screen and fixation cross both independently varied randomly in duration according to uniform distributions, between 1.2 and 1.7 s for the blank screen between 0.5 s and 0.7 s for the cross. Immediately afterwards the stimulus was presented, with its base located at the central horizontal axis of the screen. As response-locked stimulus offset might have evoked an ERP which distorts late stimulus-locked components like P3 (Busch et al., 2004) the stimulus remained on the screen for a fixed time interval of 1.2 s.3

The task was to classify the local contour (L) of the figure as a rectangle or a triangle. Participants were instructed to focus their attention on the local feature and ignore the global feature of the stimulus, and respond as fast and accurate as possible. Participants responded by pressing one of two keys using index and middle finger (counterbalanced between participants). The response pad was located on the armrest on participants’ right hand-side. We used a blocked design; in half of the blocks the irrelevant global feature was held constant (baseline condition B), resulting in two types of B blocks: B3 (with the triangle as the constant feature) and B4 (with the rectangle as constant feature). The target local feature in both B blocks varied randomly resulting in equal amounts of baseline congruent (BC) and baseline incongruent (BI) trials. In the other half of the blocks the irrelevant global and target local features were both varied randomly (filtering condition, F). This resulted in two types of F trials: filtering congruent (FC) and filtering incongruent (FI). B3, B4, and F blocks consisted of 20 trials each. Stimuli within a block were presented in random order with equal frequency and with the following restrictions: (1) no more than five consecutive stimuli were all congruent or all incongruent, (2) no more than three consecutive stimuli were identical to each other, and (3) in F blocks no more than three consecutive stimuli shared the same G feature. The first stimulus of an F block had a G feature different from the last stimulus in the preceding B block, and the first stimulus in a B block possessed the same G feature as the last stimulus in the preceding F block. Transitions between blocks occurred without warning.

A practice session of one B and one F block was carried out prior to the experiment. When the practice session started with a B condition the experiment started with an F condition (and vice versa). Practice sessions were counterbalanced with the first two blocks of the experiment. For instance, if a practice session consisted of the block sequence: [F, B3] followed by the experimental sequence [B3, F, B4, ...], then another participant had the same experimental blocks block sequence preceded by a practice session [F, B4]. One quarter of the participants started the experiment with a block sequence [B3, F, one quarter with [B4, F, one quarter with [B, F, B3], and one quarter with [F, B4]. From there on, F and B blocks alternated, with B blocks randomly chosen to be B3 or B4, with the restriction that no more than two subsequent B blocks could be of the same type. The experiment consisted of a total of 40 blocks, yielding 800 trials. All trials were presented in a single session of approximately 45 min including a 5 min break after 400 trials.

A complete set of conditions included 2 types of practice sessions (those containing B3 or B4 as their B block), 4 types of main experimental sessions (those starting with block sequences [B3, F, B4, F, [B, F, B3], or [F, B4], respectively), and 2 types of response assignments (assignment 1: index and middle finger assigned to triangular or rectangular local feature, respectively; assignment 2: vice versa). The minimal number of participants required for this counterbalanced design, was $2 \times 4 \times 2 = 16$. For each participant we evaluated performance and quality of EEG recording afterwards, one by one, excluding those with too low performance and bad recordings (see below the exclusion criteria) until the required number of participants was achieved. We discarded the data from 3 out of 19 participants; one for low performance and two for bad recording.

Electrophysiological recording

High density EEG was recorded from 61 Ag+/AgCl electrodes, placed according to the international 10/10 system using a high input

3 In a pilot behavioral study we found no difference in effects between stimulus offset at response and fixed 1.2s presentation times.
impedance amplifier (10 MΩ, BrainAmp, Brain Products GmbH, Munich, Germany) and electrodes mounted in an electrode cap (M 11, FMS, Munich, Germany). Eye blinks were monitored using two additional electrodes placed below and next to the right eye. An additional electrode was placed on the nose tip and served for offline re-referencing. Electrodes AFz and FCz were used as ground and physical reference, respectively. Electrode impedances were reduced below 5 kΩ (in 4 participants there were up to 3 channels below 10 kΩ) before data acquisition. Data were recorded with 0.53 Hz high-pass and 70 Hz low-pass filter. Data were digitized at 250 Hz.

Data analysis

The EEG-signal was filtered with a Butterworth zero-phase filter with a low cut-off frequency of 0.53 Hz (0.3 s), 24 dB/oct and high cut-off frequency of 30 Hz, 24 dB/oct. We obtained epochs of 1000 ms length including 200 ms before stimulus onset. Using a semi-automatic artifact rejection procedure, we excluded epochs if the absolute voltage difference exceeded 50 μV between two neighboring sampling points and if the amplitude was outside ±70 or −70 μV. We rejected the data from two participants because of bad recordings (more than 85% of bad epochs). Across all participants 24 (SD: 16) % of the epochs were rejected (the minimal number of trials per condition was 67). The rejected epochs were evenly distributed across four conditions BC, BI, FC, and FI. We averaged the epochs separately for each condition in each participant. We re-referenced the data to the nose electrode and corrected them to a baseline obtained from a 200 ms time window preceding stimulus onset.

Our analysis consisted of two parts. In the first part we detected and identified which of the classical ERP components showed correlates of Stroop and Garner effects. In the second part we evaluated the difference between conditions using pointwise statistical tests in each electrode, in order to detect cascaded stages from which Stroop and Garner effects arise.

In the ERP component analysis we averaged in each participant the amplitude across the electrodes within the areas selected (Fig. 2) for the ERP components P1, N1, N2, and P3. We grand-averaged data in these areas across participants and identified four ERP components in the time windows as mentioned in our Introduction. We computed the peak latencies of these components averaged over all conditions and defined the time windows for the component analysis as ±20–30 ms around the peak latency (the later the component, the larger the window). The selection resulted in the following time windows for ERP components: P1: (90–130 ms), N1: (160–200 ms), N2: (250–320 ms), and P3: (330–440 ms).

Next we computed the mean amplitude in the time window for each component in each participant. Values were compared by repeated-measures ANOVAs with two factors: Stroop (congruent vs. incongruent) and Garner (baseline vs. filtering).

We isolated the ERP correlates of Stroop and Garner effects by computing the difference between Incongruent – Congruent (Stroop) and Filtering – Baseline (Garner) conditions for each channel and also for the selected areas. To distinguish the scalp distribution of N2 and P3 activities, which partially overlap in space and time, we selected the four chains of five electrodes shown in Fig 2, located in the anterior–posterior direction in the N2–P3 areas. We averaged the difference waves across the electrodes within each chain.

---

Fig. 2. Scheme of electrode placement with outlined regions selected for the analysis of ERP components P1, N1, N2, and P3. Four chains of five electrodes used in the topographical analysis of N2 and P3 activities are marked with different colors and are numbered (1–4).

---

4 We slightly extended the upper boundary for N2 and the lower boundary for P3 to make them closer in order to depict possible N2/P3 overlap (see below).
According to this method the raw amplitude in each chain was divided by the square root of the sum of the mean squared amplitudes over the four chains. This was done for each condition and for each participant separately. A Duncan test was used for post-hoc comparison.

To compare the time courses of Stroop and Garner effects we used repeated-measures ANOVAs with two factors: Stroop (congruent vs. incongruent) and Garner (baseline vs. filtering) computed in each time point in each channel separately. We plotted the F-values as topographical maps to determine scalp regions and time course of Stroop, Garner, and their interaction. We studied the maps in each time point (4 ms at the sampling rate of 250 Hz) using pointwise ANOVAs. To avoid false positives resulting from multiple comparisons we considered effects as significant only if the difference reached p<0.05 level in at least five adjacent electrodes and for at least 11 consecutive samples (similarly to Doniger et al., 2000), i.e., if effects lasted longer than 40 ms. All significant effects in the pointwise analysis were consistent in direction and timing with the those found in the ERP component analysis.

Results and discussion

Behavioral results

We analyzed error rates and response times (RT). The overall error rate was 2.52 (SEM=0.65) %. We excluded one participant whose number of errors exceeded two standard deviations of the mean. Our RT analyses excluded error trials (n=323), and trials with RT~200 ms (n=3) or RT>1200 ms (n=30). This led to 356 (~2.78 %) exclusions from a total of 12800 trials (16 participants×800 trials). The overall mean RT was 466 (SEM=11.5 ms); condition means are given in Fig. 3. No speed-accuracy trade-off was observed.

Error rates were smaller in congruent than in incongruent trials, \( \frac{F(1,15)}{b} = 10.1, p < 0.01 \), showing both a Stroop and a Garner effect. Their interaction was marginally significant, \( \frac{F(1,15)}{b} = 4.2, p = 0.059 \).

RTs were shorter in congruent (456; SEM=11.8 ms) than in incongruent trials (476; SEM=11.5 ms), \( \frac{F(1,15)}{b} = 44.8, p < 0.001 \); and shorter in baseline (460; SEM=11.0 ms) than in filtering trials (472; SEM=12.1 ms), \( \frac{F(1,15)}{b} = 37.9, p < 0.001 \), thereby showing both Stroop and Garner effects. An interaction, \( \frac{F(1,15)}{b} = 16.0, p = 0.001 \), revealed that only the difference between filtering-congruent (458 ms) and baseline-congruent (454 ms) failed to reach significance (\( p=0.17 \) by Duncan post-hoc test). The interaction takes a super-additive form. It can be read as a stronger Stroop effect in filtering conditions, a stronger Garner effect in incongruent conditions, or both.

Stimulus repetition across trials can lead to a RT advantage (Kerns et al., 2004; Mayr et al., 2003). Baseline blocks encompass two stimuli, compared with four in filtering blocks. Smaller RTs in baseline conditions could therefore be caused by a higher proportion of trials repeating the same stimulus. We re-analyzed the RT data excluding trials in which the stimulus was the same as in the preceding trial. This analysis did not change the results: (Stroop: \( \frac{F(1,15)}{b} = 30.2, p < 0.01 \)).
p < 0.001; Garner: $F(1,15)=40.4$, $p < 0.001$; Stroop×Garner: $F(1,15)=14.5$, $p < 0.01$).

**ERP results: overview**

Table 1 contains mean amplitudes and their SEM for four ERP components corresponding to baseline-congruent, baseline-incongruent, filtering-congruent, and filtering-incongruent conditions; their averages are provided in the text.

**P1 component**
The peak latency of P1 averaged across conditions was 114 (SEM: 2.4) ms.

ANOVA of P1 amplitude revealed neither a significant Stroop, $F(1,15)=2.7$, nor a Garner effect, $F(1,15)=0.3$, nor an interaction, $F(1,15)=0.02$.

**N1 component**
The averaged peak latency of N1 was 174 (SEM: 2.9) ms. Fig. 4 illustrates the time course and topographical distribution of the N1 component for the Stroop and Garner conditions. We found a Stroop effect on N1 amplitude, which was larger in the incongruent ($-6.7$; SEM: $0.8 \mu V$) than in the congruent ($-5.9$; SEM: $0.8 \mu V$) condition, $F(1,15)=9.0$, $p < 0.01$. Neither a Garner effect, $F(1,15)=0.1$, nor an interaction, $F(1,15)=0.2$, were significant. The maximal N1 amplitude in all conditions was on the right side (at P8).

![Fig. 4. Grand-averaged (N=16) ERP, voltage maps, and difference maps and waves for the first manifestation of Stroop effect during the N1 component.](image-url)

(A) Voltage maps for the conditions: Baseline Congruent (BC), Baseline Incongruent (BI), Filtering Congruent (FC), and Filtering Incongruent (FI). The black dot denotes the N1 voltage maximum at P8. (B) Grand-averaged ERP in the N1 area (Fig. 2) for Congruent and Incongruent conditions (Stroop effect) and for Baseline and Filtering conditions (Garner effect). (C) Difference maps for the N1 time window. (D) Difference waves for the N1 area. Small bars under the curves indicate intervals where Stroop effects reached significance in pointwise tests (Fig. 8).
The averaged peak latency of N2 was 281 (SEM: 6.2) ms. In all conditions, the maximal negative amplitude of N2 was at FCz. Fig. 5 illustrates the time course and topographical distribution of the N2 component, which looks like a negative deflection on the ascending positive slope. We found a Stroop effect on N2 amplitude, which was more negative in the incongruent (3.4; SEM: 0.8 μV) than in the congruent (4.2; SEM: 0.7 μV) condition, F(1,15)=7.5, p<0.02. Neither a Garner effect, F(1,15)=1.7, nor an interaction, F(1,15)=2.3, were significant.

N2 component

The averaged peak latency of N2 was 281 (SEM: 6.2) ms. In all conditions, the maximal negative amplitude of N2 was at FCz. Fig. 5 illustrates the time course and topographical distribution of the N2 component, which looks like a negative deflection on the ascending positive slope. We found a Stroop effect on N2 amplitude, which was more negative in the incongruent (3.4; SEM: 0.8 μV) than in the congruent (4.2; SEM: 0.7 μV) condition, F(1,15)=7.5, p<0.02. Neither a Garner effect, F(1,15)=1.7, nor an interaction, F(1,15)=2.3, were significant.

P3 component

The averaged peak latency of P3 was 414 (SEM: 9.8) ms (at Cz). In all conditions, the maximal P3 amplitude was at Cz. Fig. 6 illustrates the time course and topographical distribution of the P3 component for the four conditions. We found a Garner effect on P3 amplitude, which was larger in the baseline (7.5; SEM: 1.1 μV) than in the filtering...
(7.0; SEM: 1.0 μV) condition, $F(1,15)=6.1$, $p < 0.03$. Neither a Stroop effect, $F(1,15)=1.9$, nor an interaction, $F(1,15)=3.7$, were significant.

As the N2 component was observed on the ascending slope of the P3 and these two components were located closely together, we analyzed the mean amplitude in the N2 time window at the P3 area and in the P3 time window at the N2 area (Fig. 2). We found no significant effects, nor interactions on the mean amplitude in N2 time window at P3 area; Stroop: $F(1,15)=1.9$; Garner: $F(1,15)=1.3$; Stroop×Garner: $F(1,15)=2.1$. We found a significant Garner effect on the mean amplitude in the P3 time window at the N2 area. The mean amplitude was larger in the baseline (8.1; SEM: 1.0 μV) compared to the filtering (7.5; SEM: 1.0 μV) condition, $F(1,15)=6.5$, $p < 0.03$. Neither Stroop, $F(1,15)=0.9$, nor Stroop×Garner interaction, $F(1,15)=1.2$, were significant. This observation suggests that N2 activity is more topographically delineated (local) than P3 activity. We tested the N2 and P3 topography in the anterior–posterior chains of electrodes, as described in the next section.

**Scalp distribution comparisons**

Amplitude differences for Stroop (Incongruent – Congruent) and Garner (Filtering – Baseline) are shown in Figs. 4C, D, 5C, D, 6C, D. To compare the scalp distribution of N2 and P3 components of Stroop and Garner we applied a repeated-measures ANOVA with two factors:
Location (four levels corresponding to the chains of electrodes in Fig. 2) and Interference Type (Stroop vs. Garner) on the amplitude differences. For the time window of the N2 component we found an effect of Location, $F(3,45)=6.2, p<0.02$, $\epsilon=0.43$, and an interaction, $F(3,45)=8.9, p<0.006, \epsilon=0.42$. There was no Interference Type effect, $F(1,15)=0.84$. For the time window of P3 component we found neither significant effects, nor an interaction (Location: $F(3,45)=2.5$; Interference Type: $F(1,15)<0.01$; interaction $F(3,45)=0.9$), suggesting that P3 activity was topographically widespread.

After normalization of the data using vector scaling (McCarthy and Wood, 1985) we confirmed the reliability of the Location×Interference Type interaction for the N2 time window, $F(3,45)=3.7, p<0.05$, $\epsilon=0.53$. The effect of Location after vector scaling was about significant, $F(3,45)=3.4, p<0.053, \epsilon=0.62$. The post-hoc test revealed that the difference between Stroop and Garner was significant in the first FC3–FC4 ($p<0.001$) and second C3–C4 ($p<0.05$), but not in the third CP3–CP4 ($p=0.49$) and forth P3–P4 ($p=0.65$) chain of electrodes (Fig. 7). The difference between Stroop and Garner was observed in the fronto-central electrode chains; Stroop but not Garner occurred here. Neither Stroop nor Garner effects were observed in the centro-parietal chains.

**Time course of Stroop and Garner in ERP**

We used pointwise repeated-measures ANOVAs with two factors: Stroop (congruent vs. incongruent) and Garner (baseline vs. filtering) to explore the time course of the ERP correlates of the Stroop and Garner effects. The first ERP correlate appeared 172 ms after stimulus onset; none occurred later than 400 ms after stimulus onset.\(^6\) To demonstrate the time course and topographical changes for Stroop and Garner, we plotted the time series of the amplitude difference and $F$-value maps with 20 ms step size for the interval between 160 and 440 ms after stimulus onset (Fig. 8).

Stroop reached significance in two intervals and Garner in one interval (Fig. 8B). The first Stroop interval was observed between 172–216 ms after stimulus onset and had a maximum at 200 ms, located in the parieto-occipital areas, more on the right (maximal effect at P6). The amplitude was more negative in incongruent than in congruent conditions. This interval corresponds to the negative component N1. The topographies of N1 scalp distribution (maximum at P8) and Stroop (maximum at P6) were very similar. The peak latency of N1 (174 ms) at P8 preceded the Stroop maximum (in terms of $F$-value) located at 200 ms after stimulus onset. There were no significant effects of Stroop nor Garner in the interval 220–264 ms. The second interval where Stroop reached significance occurred between 268–360 ms after stimulus onset; more negative amplitude was observed in incongruent than in congruent conditions. This interval included the negative component N2 and the rising part of the P3 component. The peak latency of N2 (281 ms) at FCz preceded the Stroop maximum (about 308 ms after stimulus onset). The scalp distribution of Stroop in the fronto-central areas (maximal effect at FC3) was shifted to the left in comparison to the FCz location of the N2 amplitude maximum.

Although slightly reduced afterwards, Stroop remained significant across the entire second interval, as can be observed in Fig. 8B, before it resurged to reach another maximum at about 336 ms. This time the effect occurred bilaterally in frontal, temporal, and parietal areas. This resurgence in Stroop overlapped with the first occurrence of Garner and Stroop–Garner interaction.

The earliest observed significant ERP correlate of the Garner effect occurred in the interval between 328–400 ms after stimulus onset. This interval corresponds to the rising part of the positive component P3. The amplitude was more negative in filtering than in baseline conditions. The peak latency of P3 averaged across conditions was observed later (414 ms) than the maximum effect of Garner at 348 ms after stimulus onset. The scalp distribution of Garner in the fronto-central areas (maximal effect at FC2) was shifted in the right anterior direction in comparison with midline maximum of P3 at Cz.

The first maximum of Garner (348 ms) nearly coincided with the second one of Stroop (336 ms). This was also the moment of maximum Stroop–Garner interaction. The interaction was observed in the interval 328–372 ms, mainly in left frontal, central, temporal, and parietal areas (Fig. 8B).\(^7\) The interaction was most prominent for the P3 electrode. For this channel we computed a pointwise Duncan post-hoc test using the same significance criterion as in the pointwise ANOVAs showed differences between all conditions except between BC–FC and BC–BI (Fig. 9). The effect sizes are very similar to those of the RT results, which showed a super-additive effect of Stroop and Garner. The present, time-resolved analysis pins these effect down to a time interval where Stroop and Garner overlap in time. They may therefore be understood as mutual reinforcement of Stroop and Garner effects.

To summarize, the Stroop effect has two early ERP correlates: one starting at 172 ms during the N1 component and the other starting at 268 ms during the N2 component. The ERP correlate of the Garner effect started later, at 328 ms, at the same time as when a resurgence in the Stroop effect takes place. At that time Stroop–Garner interaction also emerged.

**Contribution of individual shapes**

Early ERP components, such as N1, may reflect differences between individual images, independently of which Stroop or Garner condition they belong to. To see whether our congruency effects are consistent across stimuli we ran a repeated-measures ANOVA on the amplitude of N1 component with four stimuli as a single factor: G3L3, G3L4, G4L3, and G4L4. We extracted the mean amplitude of the N1 component for the electrode P6, where we observed the maximal N1 difference for Stroop effect. We found a significant effect of stimulus type, $F(3, 45)=4.34, p<0.01$. Post-hoc Duncan tests revealed that this was because of more negative amplitude in G4L4 ($-6.48, \text{SEM}=0.93 \mu V$) than in G4L1 ($-5.41, \text{SEM}=0.78 \mu V$), G3L3 ($-5.91, \text{SEM}=0.83 \mu V$) and G4L4 ($-5.86, \text{SEM}=0.95 \mu V$) were neither different

\(^6\) A Garner effect visible on $F$-value maps (Fig. 8B) at 420ms did not meet the temporal significance criterion.

\(^7\) Note: the interaction did not reach significance in the analysis of P3 component as the time window for that analysis was wide (330–440 ms), in order to embrace the whole P3, including its descending slope where the difference between conditions was not significant.
Fig. 8. (A) First row: amplitude maps for the filtering incongruent condition; second row: difference maps for Stroop effect (Congruent – Incongruent); third row: difference maps for the Garner effect (Baseline – Filtering); forth row: difference maps for Stroop – Garner. (B) F-value maps for Stroop and Garner conditions, and Stroop×Garner interaction. The scale minimum, \(F(1, 15)=4.54\), corresponds to .05 significance; colors other than dark blue over more than five electrodes reflect a significant effect. Landmark electrodes are designated on the electrode scheme located between A and B.
from G4L3 and G4L4 nor between each other. This result suggests that the N1 amplitude may reflect, besides congruency, an oblique effect. Perception of shapes composed of horizontal and vertical lines is facilitated in comparison with oblique ones (Apelle, 1972; McMahon and MacLeod, 2003). To contribute to the result, the oblique effect in the targets must be larger than that in the non-targets, presumably, because these are more salient and therefore easier to distinguish. In condition G4L4, which evoked the lowest N1 amplitude, processing is facilitated by congruence and absence of obliqueness in the target; in condition G4L3, which evoked the highest amplitude, incongruence and target obliqueness are both impeding the process. Both factors work in opposite directions in conditions G3L3 and G3L4, which have intermediate N1 amplitude. In our design the amount of horizontal–vertical and oblique segments was counterbalanced across congruent and incongruent conditions, therefore the pooled data reflect the Stroop effect only.

**ERP-behavioral correlations**

We computed the correlation between the response times, the error rates for every time point of the ERP waves across 16 participants in four conditions (BC, BI, FC, and FI). These correlations were computed for the three channels which showed the largest Stroop (P6 and FC3) and Garner (FC2). A correlation coefficient of ±0.497 was required for a 95% significance level (N = 16). Using the same restrictions for significance as with the pointwise ANOVAs, we identified the intervals and conditions where the correlation coefficients exceeded the significance level (Fig. 10).

The correlations showed similar time courses in all studied channels. For the response times, significant negative correlations were observed about 350 ms after stimulus presentation. These correlations were noticeable in all conditions but reached significance only in the congruent conditions. For the error rates significant positive correlations were observed in the interval about 400–600 ms after stimulus onset. The baseline congruent condition showed the highest correlation.

**Control experiment**

The experimental design of the main experiment contains a potentially confounding factor: baseline blocks (B3 and B4) contained only two stimuli (G3L3, G3L4 and G4L3, G4L4, respectively), but the filtering blocks contained four stimuli. Therefore it might be supposed that the higher task difficulty in the filtering compared to the baseline conditions simply resulted from the larger number of different stimuli and not from the greater variability of the irrelevant feature in the filtering condition (Pomerantz et al., 1989; Marks, 2004). We ran a control experiment, in which we equalized the number of different stimulus in the filtering and baseline conditions introducing a third feature in addition to rectangular and triangular — a half-circle (denoted O).

Four healthy, right-handed, participants (one male) aged between 24 and 40 (mean age 29.5; SD: 7.2) took part in the experiment. They were recruited among colleagues of the Leibniz Institute for Neurobiology in Magdeburg and had given consent to participate.

The experiment comprised the same stimuli as the main experiment. In order to balance the number of different stimuli between filtering and baseline blocks we included as a third feature a half-round contour shape (O) in addition to the existing triangular- and rectangular-ones. Using this feature both globally and locally, as outer and inner contour respectively, we obtained nine different stimuli (GOLO, G3L4, G4L3; G3L3, GOL4, G4LO; G4L4, GOL3 and G3LO) with which it was possible to create baseline and filtering blocks with the same probabilities of occurrence for each feature. There were altogether six types of blocks: Three types of baseline blocks (indexed by the type of constant feature: BO, B3 and B4) containing three stimuli each (BO: GOLO, GOL3, GOL4; B3: G3L3, G3LO, G3L4 and B4: G4L4, G4LO, G4L3), and three types of filtering blocks (indexed by the type of congruent stimuli: F3: G3L4, G4LO, GOL4; F4: G4L4, G3LO, GOL3; FO: GOLO, G3L4, G4L3). Note that this design balances stimulus set complexity but (necessarily) obliterates the equal probability of occurrence of congruent and incongruent stimuli. Each block was presented 7 times, yielding a total of 42 blocks, with 18 trials each and 756 trials altogether. F and B blocks were alternated and their sequence pseudo randomized, precluding the recurrence of the same block after an alternation more often than once. Within the blocks the order of stimulus presentation was also pseudo randomized with the restriction of no more than 3 identical stimuli in sequence. Participants were asked to give their response via the same response box as in the main experiment using three keys with their index, middle, and ring finger of their dominant (right) hand. Two participants were asked to use index finger for rectangular, middle finger for triangular, and ring finger for the half-circle target; one participant was asked to use ring finger for rectangular, index finger for triangular and middle finger for the half-circle target; one participant was asked to use middle finger for rectangular, ring finger for triangular and index finger for the half-circle target.

We excluded trials with errors and with RT > 200 ms (n = 17, ~ 0.55%) or RT > 1200 ms (n = 0). This led to 17 (0.6%) excluded trials out of a total of 3024 trials (4 participants ✕ 756 trials). The overall RT of the remaining trials was 571 (SEM: 37.7) ms. We calculated repeated-measures ANOVA like in the main experiment. We found effects for both main factors Stroop and Garner. Stroop: RT was shorter in congruent trials (555; SEM: 34.9 ms) as compared to incongruent trials (586; SEM: 40.1 ms), F(1,3) = 13.7, p = 0.03. Garner: RT was shorter in baseline trials (565; SEM: 36.2 ms) as compared to filtering trials (577; SEM: 39.3 ms), F(1,3) = 11.4, p = 0.04. The interaction between the factors Stroop and Garner was marginally significant, F(1,3) = 4.4, p = 0.13.

Participants had three instead of two response alternatives to choose from and used three fingers rather than two. This may explain the slower overall RT compared to the main experiment. The smaller number of participants compared to the main experiment may be responsible for failure of the Stroop and Garner interaction to reach significance. Other than this, the pattern of RT results was strikingly similar to the main experiment (Fig. 3B). Thus, equalizing the number of stimuli between the baseline and filtering conditions does not affect
the Stroop and Garner effects. In particular, it can be concluded that the Garner effect cannot simply be attributed to a larger number of different stimuli in filtering than in baseline conditions.

**General discussion**

To study the mechanisms of visual feature integration we measured Stroop and Garner effects, combining psychophysical and physiological methods. We evaluated response times, error rates, and event-related potentials. Stroop and Garner effects and their interaction were observed in RT and error rates. The ERP findings allowed us to dissociate Stroop and Garner effects both temporally and topographically. We obtained clear evidence that ERP correlates of the Stroop effect arise earlier than those of the Garner effect. Two Stroop correlates occurred, starting at 172 and 268 ms after stimulus onset, respectively, and centered above different brain regions. The only Garner correlate in ERP started appearing at 328 ms after stimulus onset. This means that there are at least two mechanisms that independently contribute to the formation of a perceptual whole; an early one that operates on the within-presentation level, presumably demarcated by approximately 100 ms inter-stimulus-intervals (Flowers, 1990) and manifested about 170 ms after stimulus onset in the ERPs, and a later one, putatively involving expectancy and memory. We observed that these mechanisms operate in cascaded stages. This implies that there is an interval where they overlap in time. Here, they seem to reinforce each other. Activity in this interval was crucially related to the response times, which may explain the frequent observation of Stroop and Garner interactions in the response times.

**Behavioral data**

The Stroop effect observed in RT is slightly unusual, as it was both obtained in baseline and filtering conditions; albeit smaller in the former than the latter. Pomerantz et al. (1989) and Marks (2004) obtained Stroop effects only in filtering, not in baseline conditions. Stroop effects in baseline conditions occurred as a result of distracter salience (Melara and Mounts, 1993). Larger salience of global distracter than local target may have played a similar role in our experiment. Using similar stimulus features (triangular and rectangular contours in one object) and task, van Leeuwen and Bakker (1995) reversed the roles of target and distracter between the global...
and local contour, which reduced the Stroop effect in the baseline condition.

Whereas Pomerantz et al. (1989) and Marks (2004) obtained a Garner effect for both congruent and incongruent stimuli; here this effect is restricted to the incongruent stimuli (although there was a similar trend in the congruent ones (Fig. 3A)). The difference in salience between target and distracter could be responsible for this asymmetry. In van Leeuwen and Bakker (1995, their Table 2), a Garner effect was observed in the congruent condition also, when the roles of target and distracter were reversed. As the pattern of the interactions was similar in RT and ERP, we will return to this issue after the discussion of the main effects in ERP.

We found correlations between individual RT and ERP around 350 ms after stimulus onset (Fig. 10), which coincides with the first manifestation of Garner and of Stroop–Garner interaction in ERP. It also coincides with N2 and P3 components and may therefore reflect cognitive control, overall task difficulty, and working memory loading. This episode, therefore, may plausibly be considered as the primary determinant of individual response speed, consistent with the idea of a central processing bottleneck. In contrast, the positive correlation of error rate with ERP in the interval about 400–600 ms indicates that errors correspond to processes operating at that late time. Given that the mean overall RT was 466 ms, this is the time where response execution takes place. Stroop and Garner effects on error rates at this stage suggest that perceptual processing conflicts lead to confusion, drops in concentration, etc., which may ultimately be responsible for motor execution failures.

**Stroop effect — first manifestation**

\( N1 \) The first ERP correlate of the Stroop effect started 172 ms after stimulus onset in the parieto-occipital areas. This latency and topography corresponded to the N1 component. N1 activity showed a higher negative amplitude for incongruent than for congruent stimuli (Fig. 4). The N1 is well-known as an index of visual spatial attention: its amplitude is higher in attentive than in inattentive conditions (reviewed in Mangun, 1995; Hillyard et al., 1998). The effect indicates that a discrimination process is applied to the attended location (Mangun and Hillyard, 1991, Vogel and Luck, 2000). N1 enhancement was found during orienting of attention to the position of a local target in a global structure in Navon-type stimuli (Proverbio et al., 1998; Han et al., 2003). In our task spatial attention similarly has to be focused in order to distinguish between alternative local targets. Larger negative amplitude of N1, therefore, could indicate that target discrimination was more difficult in incongruent than congruent conditions. Thus, Stroop conditions affect an early process taking place in visual cortical areas, whereas Garner conditions do not.

**Stroop effect — second manifestation**

\( N2 \) About 40 ms after the first ERP correlate of the Stroop effect, a second one occurred at 268 ms after stimulus onset in fronto-central areas (Figs. 5B, D, 8B). The second Stroop manifestation can be divided into two partially overlapping stages. The first coincides with the latency and topography of the N2 component; the second is synchronized with the first manifestation in ERP of the Garner effect and of Stroop–Garner interaction.

The N2 component showed higher negative amplitude for incongruent compared to congruent stimuli, similarly to the N1 component. Analyzing the Location×Condition interaction in four chains of electrodes in the N2 – P3 areas we found that the N2 activity was limited to the fronto-central areas (Fig. 7). This suggests that the Stroop effect reflected in N2 activity originates from the medial frontal cortex, possibly including the anterior cingulate (Ridderinkhof et al., 2004).

The observed fronto-central N2 is most likely one of the anterior subcomponents related to mismatch detection or to cognitive control, as discussed by Folstein and Van Petten (2008). In our paradigm target and non-target features were balanced (unlike the oddball paradigm typical for mismatch N2), therefore it is unlikely that the observed N2 is related to mismatch detection; more likely, it is related to cognitive control processes involved in conflict detection and monitoring. These may result in larger N2 amplitude in incongruent than in congruent conditions.

N2 amplitude, both in auditory (Nieuwenhuis et al., 2004) and visual tasks (Folstein and Van Petten, 2004; Azizian et al., 2006), depends on perceptual overlap between targets and non-targets which leads to the recruitment of the N2-related control processes. In our stimuli, the integration of the target and non-target features results in the emergence of a perceptual whole: a closed shape. Hence there is substantial perceptual overlap. Lack of perceptual overlap may explain the absence of N2 modulation in previous ERP studies using the classical Stroop task (Rebai et al., 1997; Lew et al., 1997; West and Alain, 1999; Liotti et al., 2000). Some of these studies found larger fronto-central negativity in incongruent than congruent stimuli (Rebai et al., 1997; West and Alain, 1999; Liotti et al., 2000), but this negativity was observed later, about 450 ms after the stimulus presentation. As argued, late onset negativity in the classical Stroop task may reflect the time needed for detection of conflict between perceptual and semantic information streams.

Early during perception only within-presentation information leads to detectable interference in the ERP; still at peak latency of N2 (281 ms) only Stroop congruency has been observed. However, very soon afterwards, at 328 ms, the Garner effect appears: from this point we start to observe interference from information at the between-presentations level.

**Garner effect — first manifestation**

\( P3 \) The first ERP correlate of a Garner effect appeared in our study at 328 ms, lasting until 400 ms after stimulus onset. This effect corresponded to the rising part of the positive component P3 but did not last until its peak latency, at 414 ms (Figs. 6B, D, 8B). As in previous ERP studies of Garner interference (Lew et al., 1997; Kaganovich et al., 2006; Caclin et al., 2008), the amplitude was lower in filtering than in baseline conditions.

P3 amplitude is classically understood to reflect the update of mental representation by new information — “context updating” (Donchin, 1981); the lower a target's likelihood the larger P3. However, numerous studies evidenced that P3 amplitude also depends on factors such as task relevance (i.e., amount of attention devoted to the stimulus) and task difficulty (reviewed in Kok, 2001, Polich, 2007). In addition, P3 amplitude may be affected by memory processes which are reflected in negative slow wave activity overlapping with P3 (Wijers et al., 1989). As Garner interference occurs between trials, it involves memory. Whereas low target probability and high task relevance lead to larger P3, task difficulty and memory load have the opposite effect. The reduced P3 amplitude in the filtering condition therefore may indicate that increased task difficulty and memory load due to non-target variability across trials (Pomerantz et al., 1989; Marks, 2004) overrides any positive effects of the task-irrelevant stimulus variability.

Besides cognitive control, is also memory involved? The Garner effect in the rising part of P3 overlapped with N2 (Fig. 6B) and also was present in the fronto-central areas associated with N2. This suggests that the larger P3 negativity in the filtering condition may be intensified by the N2-related processes of cognitive control. The combination of N2 and P3 activities is commonly observed, and peak-to-peak N2–P3 amplitude is even used as a measure of activity of the frontal networks (e.g., Daffner et al., 2000). In Garner studies, N2–P3 overlap was proposed to explain lower P3 amplitude in filtering than...
Caclin et al. (2008) observed that the stimuli used by Kaganovich et al. (2006) and Caclin et al. (2008) were not optimally suitable to be used as a tool to understand attentional processing. In contrast, with similar experimental designs for which Stroop without Garner interference was found earlier than in our study. Of these, the results of Lew et al. (1997), where classical Stroop stimuli were used, and therefore the semantic and perceptual domains were mixed, can hardly be compared to ours. The two other studies were done within the auditory perceptual domain. Auditory processes, as reflected in ERP components, are generally faster than visual ones. Kaganovich et al. (2006) asked their participants to discriminate between two vowels or two male talkers in baseline, filtering, and correlated conditions. Garner interference was observed in sustained negativity which, similarly to our study, was most prominent at N2 and P3 latencies, about 200–350 ms after stimulus onset, but started from the auditory N1 component with about 100 ms latency. The authors explained the sustained effect as a result of the generally high attentional effort required for the filtering task. Caclin et al. (2008) proposed a similar explanation for the very early Garner effect (at about 30 ms after stimulus) which they observed in one of two groups. This group received timbre dimensions, one relevant and the other irrelevant, which were highly overlapping.

It could be argued that the particular dimensions investigated determined the size of the filtering costs and therefore probably determined the presence or absence of early attentional effects observed in the ERPs. On the other hand, it is likely that identifying the target in conditions where perceptually highly similar distracters were randomly varied is difficult. Hence, in compensation, extra resources will be deployed to the early processing stage filtering conditions, which would explain the early effects. This means that the stimuli used by Kaganovich et al. (2006) and Caclin et al. (2008) were not optimally suitable to be used as a tool to understand feature integration. In contrasts with Kaganovich et al. (2006) and Caclin et al. (2008), our stimuli did not require excessive attentional effort for the filtering task. Caclin et al. (2008) noted that no filtering was detectable in the filtering condition.

The present results support the distinction of underlying mechanisms for Stroop and Garner effects. These mechanisms operate at different time moments, according to their ERP correlates and, according to their scalp distributions, involve different brain areas. The conflict responsible for the Stroop effect started from the “visual” occipital areas, and then, after some delay, appeared in the “control”-related fronto-central areas. Thus, the Stroop effect includes both visual and cognitive control conflicts. In contrast, for the Garner effect no visual component was detectable in the ERPs.

Stroop and Garner effects overlapped only for 40 ms. But even in this overlapping interval the topography of the effects was different: a fronto-central maximum for Garner and lateral maxima for Stroop (Fig. 8B). In the parietal areas, the Garner effect lasted about 40 ms after the end of all Stroop effects. The difference in topography confirms the distinction in mechanisms.

At the onset of the Garner effect in ERP at 328 ms, an interaction between Stroop and Garner also started (Figs. 8B, 9). The ensuing time interval is likely to be the main determinant of individual RT. The interaction is super-additive and very similar to that of RT (Fig. 3A). Correlations between RT and ERP waves (Fig. 10) indicate that the interaction in RT originates during this time interval. In Sternberg’s (1969) additive factors logic, the interaction in RT between Stroop and Garner effects excludes that the two belong to different sequential stages. Time-resolved electrophysiology shows these effects to belong to cascaded stages. The short period of overlap between the stages is sufficient for them to mutually reinforce their effects on the response times, because this period is the main determinant of RT. We may consider it a central bottleneck for Stroop and Garner-related processes.

The interaction observed here and in the RT differs from that amply debated in the psychophysical literature. Pomerantz et al. (1989) and Marks (2004) observed that Stroop effects were found only in the condition where the irrelevant feature varied (filtering), not within the blocks in which it was kept constant (baseline). A Garner effect was found in both congruent and incongruent conditions. The authors proposed that leakage between processing channels may be the common source of Stroop and Garner effects. As Stroop in addition requires incongruence of the information processed between channels, “Stroop without Garner” would become unlikely. Pomerantz et al. (1989) reported that despite much effort, they failed in creating an experimental design for which Stroop without Garner interference could be observed. They concluded that Garner effects were more robust and therefore a more sensitive measure of feature integration than Stroop effects.

This misleading conclusion may have resulted from an unfortunate choice of stimuli. For instance, Pomerantz (1983) famously used pairs of parentheses “)” “(”, “)”, “(”, and “). The task was to classify the orientation “)” “)” “)” “(” “)(” one of the parentheses of the pair and ignore the other. The first two pairs are same in orientation and thus were considered to be congruent, whereas the other two pairs were considered incongruent. Stroop effects were not reliably observed in this case. One problem with these stimuli, however, is that they are graphemic symbols (cf. Garner, 1988). In particular, letters show different patterns of interference than geometrical shapes. Letters often show no Stroop effect (Lachmann and van Leeuwen, 2004) or even a negative one — faster and more accurate responses in incongruent than in congruent conditions (van Leeuwen and Lachmann, 2004), while corresponding non-letter shapes show a robust Stroop effect. In our understanding, these results point to an acquired early processing distinction between graphical and geometrical stimuli relevant to reading and the problems in learning to read (Lachmann and van Leeuwen, 2008).

Several studies have subsequently shown that Stroop without corresponding Garner effect could be obtained (Melara and Mounts, 1993; van Leeuwen and Bakker, 1995; Patching and Quinlan, 2002). Combined with Pomerantz’ observations on Garner without Stroop, this constitutes a double dissociation in effect. Although far from fail-safe (Van Orden et al., 2001), such results are used as a diagnostic to infer that two effects belong to different processing mechanisms. Our present study confirms this conclusion for Stroop and Garner effects.

Conclusion

Target and distracter features belonging to the contour of a powerful, closed shape resulted in strong and robust Stroop and Garner effects. We confirmed that these effects rely on different mechanisms that belong to cascaded processing stages. The Stroop effect is manifested earlier in time course of perceptual organization.
and is relevant to the mechanisms of attentional discrimination and cognitive control. The Garner effect encompasses a mixture of processes related to cognitive control, task difficulty, and working memory. Where Stroop and Garner processes overlap in time, they reinforce each other. As this happens in a time interval which is the primary determinant of processing time, Stroop and Garner, in spite of their independence, are generally found to interact in response times.

Acknowledgments

Part of this work was supported by a grant from the European IST Programme Project FP6-002778 (“DIRAC”). We thank two anonymous reviewers whose valuable suggestions helped to improve earlier versions of our manuscript.

References

Dishon-Berkovits, M., Algom, D., 2000. The Stroop effect: it is not the robust phenomenon that you have thought it to be. Mem. Cognit. 28, 1437–1449.


