Received: 14 May 2020 | Revised: 12 November 2020 | Accepted: 16 November 2020

DOI: 10.1111/psyp.13744

ORIGINAL ARTICLE

PSYCHOPHYSIOLOGY SPR WILEY

Which digit is larger? Brain responses to number and size interactions in a numerical Stroop task

Hsu-Wen Huang[1] | **Mauro Nascimben**[2] | **Ya-Yi Wang**[3] | **Dong-Yang Fong**[4] | **Ovid J.-L. Tzeng**[5,6,7,8] | **Chih-Mao Huang**[2,8]

[1]Department of Linguistics and Translation, City University of Hong Kong, Kowloon, Hong Kong

[2]Department of Biological Science and Technology, National Chiao Tung University, Hsinchu, Taiwan

[3]Cognitive Neuroscience Laboratory, Institute of Linguistics Academia Sinica, Taipei, Taiwan

[4]Physical Education Office, National Taipei University of Technology, Taipei, Taiwan

[5]College of Humanities and Social Sciences, Taipei Medical University, Taipei, Taiwan

[6]Department of Educational Psychology and Counseling, National Taiwan Normal University, Taipei, Taiwan

[7]Hong Kong Institute for Advanced Study, City University of Hong Kong, Kowloon, Hong Kong

[8]Center for Intelligent Drug Systems and Smart Bio-devices (IDS^{2}B), National Chiao Tung University, Hsinchu, Taiwan

Correspondence
Chih-Mao Huang, Department of Biological Science and Technology, National Chiao Tung University, Hsinchu, Taiwan.
Email: cmhuang@nctu.edu.tw

Funding information
Academia Sinica Thematic Research Program, Grant/Award Number: AS-103-TP-C04; Taiwan's Ministry of Science and Technology, Grant/Award Number: 105-2420-H-009-001-MY2, 106-2410-H-001-024-MY2, 107-2410-H-009-028-MY3 and 108-2321-B-038-005-MY2; Center for Intelligent Drug Systems and Smart Bio-devices; Higher Education Sprout Project of the Ministry of Education in Taiwan; Hong Kong Institute for Advanced Study, Grant/Award Number: 9360157; City University of Hong Kong, Grant/Award Number: 7200538

Abstract

When comparing the digits of different physical sizes, the processing of numerical value interacts with the processing of physical size. Given the universal use of Arabic numbers in mathematics and daily life, this study aims to elucidate the cognitive processes involved in the interactions of task-relevant and task-irrelevant features during information processing. We investigated this question by examining event-related potential (ERP) using a modified version of the size congruity comparison, which is a Stroop-like task. Numerical value and physical size were varied independently under task-relevant and task-irrelevant conditions. To better examine how the task-irrelevant features modulated the processing of the task-relevant attributes, a neutral condition was included in both tasks. For the physical task, congruent trials showed a less negative N200 response than neutral trials (indicating a facilitation effect), and incongruent trials elicited a larger N450 and smaller late positive complex (LPC) response than neutral trials (indicating an interference effect). For the numerical task, congruent trials showed a larger LPC response than neutral trials (indicating a facilitation effect). These ERP findings indicate that the sources of the facilitation and interference effects appear in different cognitive processes for each task. We further suggest that language characteristics may be a factor in the superior numerical processing exhibited in this study.

KEYWORDS

ERP, facilitation, interference, LPC, N200, N450, numerical Stroop

1 | INTRODUCTION

"Automatic processing" refers to the mental ability to rapidly and efficiently process information, and depends on the importance of the information or proficiency in the task (Posner & Snyder, 2004; Schneider & Shiffrin, 1977). Examples of automatic processing are the perception and evaluation of physical size (discrete or non-symbolic quantities) and some studies also suggested that the processing of numbers (abstract or symbolic quantities) is automatic (Dehaene, 1997; but see Cohen Kadosh & Walsh, 2009 for a discussion of this issue). Despite there are different views on whether number processing is automatic, numerous studies have reported that the parietal cortex is a common site for the cerebral representation of quantity (Cohen Kadosh et al., 2005; Cohen Kadosh & Henik, 2006; Dehaene et al., 1998; Fias et al., 2003; Huang et al., 2012). However, how the above two types of quantity interact during information processing remains unclear (for a review, see Henik et al., 2017). Some models suggest that size and number interact by sharing a common magnitude-representation process (e.g., Bueti & Walsh, 2009; Walsh, 2003), whereas other models suggest that size and numbers are processed separately, but interact at the decision level (e.g., Schwarz & Heinze, 1998), although experimental support for these ideas has yet to be fully established.

The size congruity effect, also referred to as the numerical Stroop effect (Besner & Coltheart, 1979; Henik & Tzelgov, 1982), plays a pivotal role in this discussion. In the study that established the original size congruity paradigm (Besner & Coltheart, 1979), the participants compared the numerical values of two simultaneously presented digits that differed in both numerical magnitude (task-relevant) and physical size (task-irrelevant). Three conditions were used: in the congruent condition, the numerically greater number was also physically larger (e.g., "3 8"); in the neutral condition, only the numerical magnitude varied between the two digits, with the physical sizes equal (e.g., "3 8"); and in the incongruent condition, the numerically greater number was physically smaller (e.g., "3 8"). Based on the size congruity effect, response times were shorter in the congruent condition and longer in the incongruent condition. These findings suggested that numbers are intimately associated with size, such that the physical size of a digit modulates the numerical comparison response time. The results were further interpreted as supporting a simple serial processing model, in which physical size is processed first and then, influences the subsequent number processing (but not vice versa).

However, evidence against this serial account was obtained by Henik and Tzelgov (1982), who utilized a modified numerical Stroop paradigm with both a physical size judgment task and a numerical magnitude judgment task. To replicate the congruency effect in the numerical task, the authors further demonstrated the congruency effect in the physical task (Henik & Tzelgov, 1982). In both tasks, compared with the neutral conditions, the subjects exhibited a facilitation effect (with shorter reaction times and/or higher accuracy) in the congruent condition and an interference effect (with longer reaction times and/or lower accuracy) in the incongruent condition. These findings suggested that physical size and numerical magnitude influence each other bi-directionally, regardless of whether they represent task-relevant or task-irrelevant information. The two stimulus attributes are processed automatically and in parallel, but not completely independently (MacLeod, 1991).

Two main models have been proposed for the processing interactions between task-relevant and task-irrelevant features in the numerical Stroop paradigm. One view posits that numerical and physical sizes are mapped onto an amodal representation (e.g., Bueti & Walsh, 2009; Walsh, 2003). According to this account, the congruency effect originates at the stimulus representational level before task-specific responses occur; it has thus been called the early interaction account (Schwarz & Heinze, 1998). Other views, such as the late interaction account or dual-route model (Santens & Verguts, 2011; Schwarz & Heinze, 1998), posit that numerical size and physical size are initially processed in parallel and separately, but compete to dominate response activity at the decision level. Late interaction accounts thus posit a response conflict based on activity in different systems, whereas the early interaction account posits a stimulus conflict within a single system.

Several studies using functional magnetic resonance imaging (fMRI) have demonstrated that the congruity of the physical size and numerical value modulates the intraparietal sulcus activation in the numerical Stroop paradigm, with the incongruent condition yielding greater activation than the congruent condition (Kaufmann et al., 2005; Pinel et al., 2004). Along with other research showing that the intraparietal sulcus is responsive when comparing physical size and numerical value (Cohen Kadosh & Henik, 2006; Cohen Kadosh et al., 2005), this result was used to support the amodal representation, that is, the early interaction account (Szűcs & Soltész, 2008). However, some studies have also observed greater activation in several fronto-parietal regions for a numerical magnitude task than for a physical size task (Pinel et al., 2001, 2004), but no further explanation has been given for this difference. Furthermore, Cohen Kadosh et al. (2007) revealed interference effects in the primary motor cortex, and hence suggested that the conflict between numerical and physical magnitude is not completely resolved until response initiation. To fully understand when and how task-relevant and task-irrelevant features interact in the physical size and numerical magnitude tasks of the numerical Stroop paradigm, time and timing must be considered.

Event-related potentials (ERPs) offer a direct, temporally precise, multidimensional view of brain activity, including functionally specific neural markers for aspects of perception, attention, memory, and language (Fabiani et al., 2007). This makes them particularly well suited for exploring the neural underpinnings of physical and numerical magnitude interactions. Different ERP signatures have been associated with conflict monitoring and stimulus evaluation. In particular, cognitive control has been associated with a frontal negativity appearing at around 200–350 ms, known as the N200 (Folstein & Van Petten, 2008), while differences in the ease of stimulus evaluation can be observed as modulations of the widely distributed late positive component appearing at around 300–600 ms and known as the late positive complex (LPC) or P300 (Coles & Rugg, 1995; Kok, 2001; Verleger et al., 2005).

To examine the locus of the congruity effect under the numerical Stroop paradigm, Schwarz and Heinze (1998) compared single digits with a fixed numerical or physical standard of 5 in an ERP study. Six active electrodes were recorded and referenced to the left ear lobe. In the numerical task, a congruity effect was observed on FZ between 368 and 400 ms as a positive-going waveform, with the congruent conditions more positive than the incongruent conditions. In the physical task, the congruity effect was observed on the N200 (280–420 ms) from FZ and on the P300 from PZ, with incongruent conditions eliciting a more negative N200 response and a later P300 response than congruent conditions. The authors interpreted the congruency effects found on FZ as consistent with the early interaction account. However, the cognitive processes involved in the two tasks were not clearly discussed, which is possibly due to the different ERP patterns observed on FZ. Research has shown that the anterior N200, which requires attention to the eliciting stimulus, is always accompanied by P300 (Folstein & Van Petten, 2008). Examination of fig. 2 in Schwarz and Heinze (1998) shows that the frontal N200 in the numerical task overlaps with subsequent positivity; thus it seems the congruency effect is evident on the beginning of the frontal P300 instead of on the N200, which in turn suggests that the task-irrelevant physical size modulates the response processing. In contrast, the congruency effect on the frontal N200 in the physical task is clear, which indicates that the task-irrelevant number modulates the stimulus identification process. Therefore, there is insufficient evidence to support the early interaction hypothesis.

In another ERP study, Cohen Kadosh and colleagues used the modified numerical Stroop paradigm for both physical and numerical tasks (Cohen Kadosh et al., 2007). The ERP data were recorded from 128 electrodes and referenced to the CZ. The congruency effects for both tasks were observed on the P300; therefore, the authors suggested that their results supported the early interaction account, according to which the conflict originates from the perceptual level. However, this is questionable, because the P300 is associated with stimulus evaluation, which happens after perceptual processing (Donchin, 1981). Furthermore, the study presented only ERP data from one electrode (PZ), providing no information on the possible frontal effect, such as the N200. Overall, the ERP evidence is still inconclusive regarding how physical size and numerical number interact in the numerical Stroop paradigm.

The disparity across studies suggests that multiple mechanisms may underlie behavioral congruency effects, with stimulus- and task-related factors determining which set of mechanisms are engaged (e.g., Dadon & Henik, 2017). On top of the experimental factors, linguistic properties, culture, and mathematics learning strategies may also affect numerical magnitude representation and performance (Miura & Okamoto, 2003; Nuerk et al., 2005; Pixner et al., 2011). For example, the Chinese language is highly transparent with respect to the power of a given number (e.g., "ten one" for 11), and the phonological structure of Chinese number names is much simpler than in English. In Chinese, all single-digit numbers have single syllables, whereas in English the name for "7" contains two syllables. Studies have demonstrated that these linguistic properties facilitate Arabic number comparisons, counting, and general numerical performance using Arabic notation for Chinese speakers (Miura & Okamoto, 2003). The literature examining interactions between numbers and size suggests that among native English speakers, numerical values are processed later than physical size for both numerical and physical tasks (Schwarz & Ischebeck, 2003; Szűcs & Soltész, 2007). Given the universal use of Arabic numbers in mathematics and daily life, this study aims to elucidate the cognitive processes involved in the task-relevant and task-irrelevant features of a numerical Stroop task for native Chinese speakers.

The two dimensions (physical size and numerical magnitude) are either congruent or incongruent with each other, corresponding to the congruent and incongruent conditions in this study. Additionally, there is a third possible relationship between the two dimensions, in which one dimension has the same value while the other varies; this corresponds to the neutral condition. Having a neutral condition as a baseline, we can further examine how facilitation and interference effects contribute to the early and late stages of information processing. If the task-relevant and task-irrelevant attributes interact at the early stages of information processing, we expect to see the facilitation and/or interference effects on the N200, and this finding would support the early interaction account. Moreover, if the task-relevant and task-irrelevant attributes compete to dominate response activity at the decision level (late interaction account), we expect to see the facilitation and/or interference effects on the P300. Furthermore, if numerical values are processed later than the physical size as suggested in the literature for both numerical and physical tasks (Schwarz & Ischebeck, 2003; Szűcs & Soltész, 2007), then task-irrelevant numerical value would affect physical size judgment at the later stage of processing. And, the task-irrelevant physical size would modulate numerical judgment at the earlier stage of processing.

2 | MATERIALS AND METHOD

2.1 | Participants

Twenty healthy, right-handed young adults between the ages of 20 and 26 (12 females; mean age 23.5 years) participated in this study. All of the participants were native Chinese speakers with no early English exposure before age 12. However, all of them would have learned English from 7th to 12th grade in school because of the language education policy of Taiwan. The participants have filled in a language questionnaire indicating whether they use (read, write, and speak) English in their daily life and their general English proficiency level. Most of the participants (18 out of 20) reported that they do not use English in their daily life. Only two participants reported that they occasionally use English, and their overall English proficiency level is low-intermediate. The participants provided written informed consent prior to their participation and the study was approved by the Institutional Review Board of Academia Sinica, Taiwan. The participants were screened for normal vision and had no history of neurological or psychiatric disorders.

2.2 | Stimuli

We used a modified version of the physical-numerical interference paradigm (Henik & Tzelgov, 1982). The stimuli were pairs of Arabic digits shown simultaneously in the middle of a computer screen. Two types of magnitude judgment were included in the experiment: a numerical magnitude task and a physical size task. In the numerical magnitude task, the participants were asked to judge which digit was numerically larger, while ignoring the physical size of the digits on the screen. In the physical size task, the participants were asked to judge which digit was physically larger, while ignoring the numerical magnitude of the digits. For both tasks, individual digits between 1 and 9, but excluding 5 (Tzelgov et al., 1992), were used to create the digit pairs.

For each task, the digit pairs were created with three levels of congruity. In congruent trials, the numerically larger digit was also physically larger. In incongruent trials, the numerically larger digit was physically smaller. In neutral trials, we matched the task-irrelevant dimension but made the task-relevant dimension different: for physical neutral trials, the digit pairs had the same numerical value; for numerical neutral trials, the digit pairs had the same physical size. The digit pairs were presented in Arial font at two different sizes (55 and 73) as required to manipulate the physical size of the stimuli. The stimuli in the neutral condition of the numerical magnitude task were presented at a medium font size of 64. The large and small numbers (physically or numerically) were evenly distributed across the left and right sides.

There were two blocks of 48 trials for each task, yielding 192 trials in total. Within each block, one third of the trials were conducted in each of the congruent, incongruent, and neutral conditions. The experimental stimuli were presented via MATLAB (MathWorks, version 2009, 32 bit) Psychtoolbox (version 2.54). The order of the two tasks was counterbalanced, and the stimuli were randomly presented.

2.3 | Procedure

At the beginning of each block, the participants were visually instructed to identify which digit in the digit pair was numerically larger or physically larger, and to respond as quickly and accurately as possible with a judgment by pressing a button. They were asked to press the left mouse button if they deemed that the left digit was numerically larger (in the numerical magnitude task) or physically larger (in the physical size task), or to press the right mouse button if the right digit was larger (in the numerical magnitude task) or physically larger (in the physical size task). For each trial, a white cross appeared centrally for 500 ms followed by a blank screen for a variable inter-stimulus time interval (ISI; a variable interval was used to temporally jitter anticipatory ERP responses) of between 300 and 800 ms. Next, a digit pair was presented on the screen for a maximum of 1,000 ms, or until the participant responded, followed by a blank screen for a minimum of 200 ms. The participants were encouraged to minimize blinks or eye movements during this period. In total, the participants had a maximum of 1,200 ms to respond by pressing a mouse button. The index finger of the right hand was positioned over the left button while the middle finger was over the right mouse button. At the end of each trial, a capital B was displayed for 1,500 ms, indicating that blinking was now allowed. The inter-trial interval was 500 ms. There were four blocks of trials, with 48 trials per block. The order of the blocks was presented in an ABBA sequence for half of the participants and a BAAB sequence for the other half. Between blocks, the participants took a short break.

2.4 | EEG recording and processing

Each electroencephalogram (EEG) was recorded from 32 Ag/AgCl electrodes mounted on a cap (QuikCap, Neuromedical Supplies, Sterling, VA, United States). Signals were amplified by a SYNAMPS2 device (Neuroscan, Inc.) with

0.1–100 Hz band-pass and digitized at 500 Hz. Data were referenced to the average of the left and right mastoids. Eye movement was monitored using two electrodes attached to the supra-outer canthus of the left eye and infra-outer canthus of the right eye.

The EEG data were preprocessed in the EEGLAB environment (Delorme & Makeig, 2004). Data from 20 healthy subjects were visually inspected and cleaned using the Icaeyeblinkmetrics toolbox (Pontifex et al., 2017) and the artifact subspace reconstruction method (Chang et al., 2018). Initially, we used independent components to remove eye blinks based on vertical electrooculograms as a reference template. As parameters for blink removal, we used a correlation threshold of 0.9 between template and signal for eye blink identification and 0.8 as the minimal correlation between artifact and signal. After blink rejection, we used artifact subspace recognition to handle both transient and large amplitude artifacts, depending on signal variance. EEG traces were then band-pass filtered in the range 0.2–30 Hz using a FIR filter with Hamming window and filter order calculated according to signal length. ERPs were then computed from a 500 ms prestimulus baseline to a 1,200 ms poststimulus onset. If the signals exceeded five standard deviations in amplitude, the whole epoch was rejected.

3 | RESULTS

3.1 | Behavioral results

For each task, the data on reaction time and accuracy were subjected to repeated-measures analyses of variance (ANOVAs) with the three levels of congruity as within-subject factors. Incorrect responses or reaction times that exceeded two standard deviations from an individual subject's mean were removed from the analysis.

TABLE 1 Mean reaction times and accuracy

Tasks	Condition	Reaction time (ms)	Accuracy (%)
Numerical	Congruent	538.3 (79.3)	99.7 (1.4)
	Neutral	573.3 (85.5)	98.9 (2.1)
	Incongruent	619.8 (78.0)	93.1 (5.9)
Physical	Congruent	461.5 (59.2)	99.7 (1.0)
	Neutral	470.1 (62.4)	99.8 (0.7)
	Incongruent	513.3 (82.0)	96.4 (3.7)

For reaction times, the congruity effect was significant for both the numerical task [$F(1, 19) = 56.3$, $p < .001$] and the physical task [$F(1, 19) = 29.0$, $p < .001$]. Pair-wise comparisons showed that both facilitation (where the congruent condition had a shorter reaction time than the neutral condition) and interference effects (where the incongruent condition had a longer reaction time than the neutral condition) were significant in the numerical task ($p < .001$). In the physical task, only the interference effect was significant ($p < .001$). For accuracy, the congruity effect was significant for both the numerical task [$F(1, 19) = 23.3$, $p < .001$] and physical task [$F(1, 19) = 15.5$, $p < .001$]. Post hoc analyses showed that the interference effect was significant for both tasks ($ps < .001$), with the incongruent condition having a lower accuracy than the neutral condition (Table 1).

3.2 | ERP results

Figures 1 and 2 overlay the grand average ERPs at five representative channels for the congruent, neutral, and incongruent conditions in the numerical task and the physical task, respectively. Only the correctly answered trials were included in the ERP plots and data analysis. All conditions elicited typical brain responses for visual stimulation,

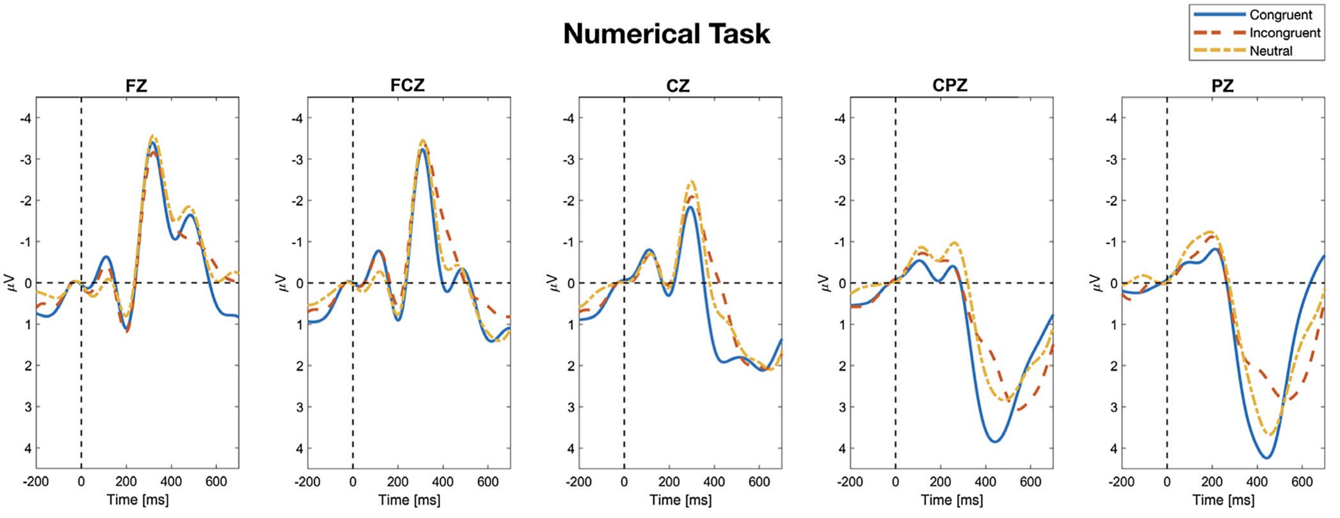

FIGURE 1 Grand average ERPs of the congruent, incongruent, and neutral trials for the numerical task

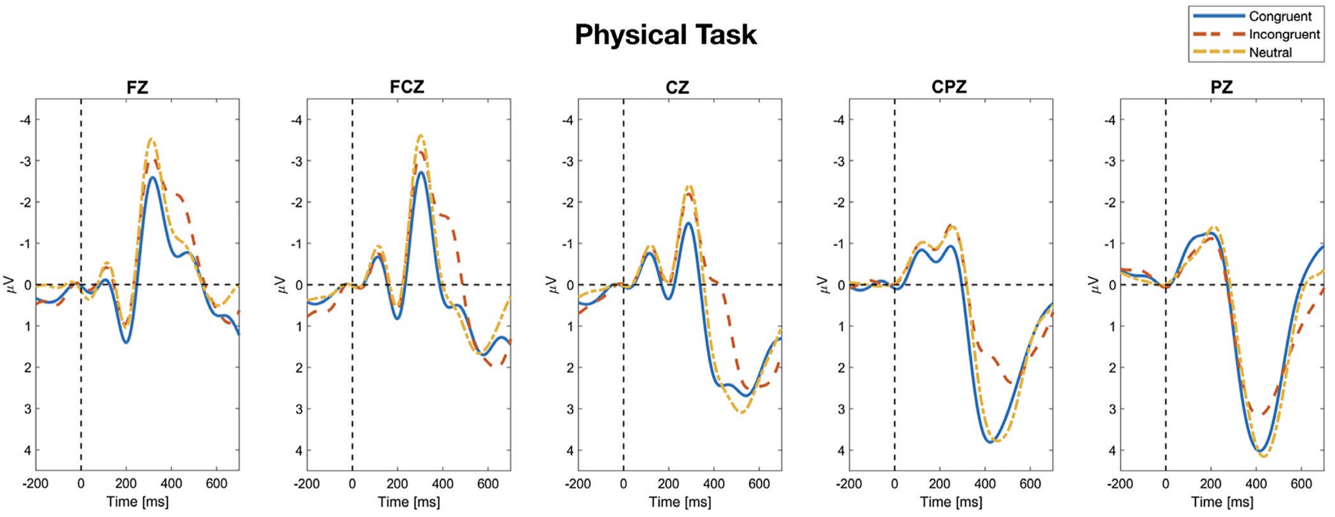

FIGURE 2 Grand average ERPs of the congruent, incongruent, and neutral trials for the physical task

including the posterior P1 and N1, and the anterior N1 and P2. Following the sensory components, all conditions elicited a negative-going wave (N200) and a second negativity (N450) at frontal-central sites, and then, a parietal-distributed broad LPC. For each task, the congruency effects were analyzed for the anterior N200 (200–380 ms), anterior N450 (380–520 ms), early LPC (320–520 ms), and late LPC (520–720 ms) by performing ANOVAs with the three levels of congruency (congruent, neutral, and incongruent), and with electrodes in the regions of interest. For each ANOVA, the Greenhouse–Geisser adjustment to the degrees of freedom was applied to correct for violations of sphericity associated with repeated measures. For all F tests with more than one degree of freedom in the numerator, the corrected p value is reported. In the analysis of the N200 and N450, congruency and electrode (FZ, F3/4, FCZ, and FC3/4) were included as within-subject factors. In the analysis of the early and late LPCs, congruency and electrode (CPZ, CP3/4, PZ, and P3/4) were included as within-subject factors.

3.2.1 | Numerical task

The congruency effects were *ns* on either the N200 or the N450 window (Fs < 1). There was a main effect of congruency on the early LPC [$F(2, 38) = 9.5$, $p < .001$], showing both a facilitation effect (a more positive response in congruent trials than in neutral trials) and a congruency effect (a more positive response in congruent trials than in incongruent trials) in the pair-wise comparisons. The congruency main effect on the late LPC was also significant [$F(2, 38) = 6.3$, $p < .01$], with pair-wise analyses showing that only the congruency effect (congruent vs. incongruent) was significant.

3.2.2 | Physical task

There was a significant congruency main effect on the N200 [$F(2, 38) = 5.6$, $p < .05$], with pair-wise analyses showing that congruent trials elicited a less negative N200 than neutral trials (indicating a facilitation effect) and incongruent trials (indicating a congruency effect). The congruency main effect was significant on the N450 [$F(2, 38) = 5.4$, $p < .05$], with pair-wise analyses showing that the incongruent condition elicited more negative responses than the neutral condition (indicating an interference effect) and the congruent condition (indicating a congruency effect). On the early LPC, a congruency main effect was observed [$F(2,38) = 9.3$, $p < .01$], with pair-wise comparisons showing that the incongruent trials elicited a less positive response than the neutral trials (indicating an interference effect) and congruent trials (indicating a congruency effect). No congruency main effect was found on the late LPC [$F < 1$].

3.3 | The relationship between LPC and reaction time

Previous research has suggested that the LPC reflects stimulus evaluation (Coles & Rugg, 1995) and the attentional resources required for a task (Polich, 2007). To examine the relationship between LPC mean amplitude, peak latency, and reaction times, a multiple linear regression was calculated to predict median reaction times based on the LPC mean amplitude and peak latency (Figure 3). The results provide information on whether the LPC mean amplitude or timing was primarily associated with behavioral response speed. Congruency may affect the degree of latency jitter of the LPC: LPC amplitudes are greater (smaller) and latencies are

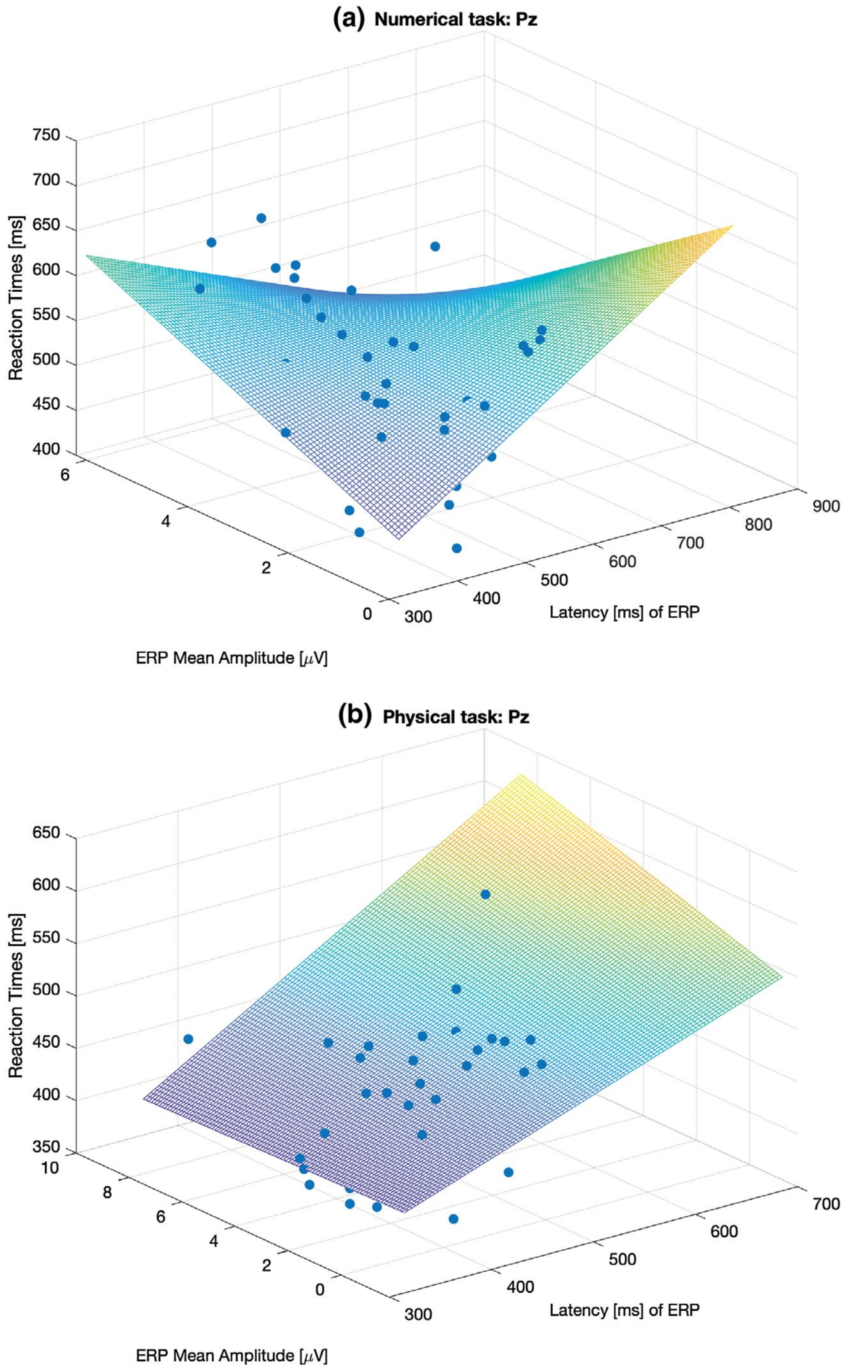

FIGURE 3 Multiple regression between LPC mean amplitude, LPC peak latency, and reaction times for the numerical task (a) and the physical task (b)

shorter (longer) for congruent (incongruent) trials. Therefore, a time window of 320–800 ms was used to better capture the broad positivity. Multiple regression equations were reliable for both the numerical task ($R^2 = .29$, $p < .001$) and the physical task ($R^2 = .28$, $p < .001$). For the physical task, the LPC peak latency was primarily associated with behavioral response time.

As depicted in Figure 4, the single regression correlations of LPC peak latency on Pz and reaction times for the numerical task and physical task were $R = .34$ ($R^2 = .12$, $p < .01$) and $R = .52$ ($R^2 = .27$, $p < .001$), respectively.

3.4 | EMS spatial filter

We applied a spatial filter (Schurger et al., 2013) to further validate our data. This method is blind to prior electrode selection, avoiding regional bias when deciding on the time-window (component) analysis. Pre-selecting a subset of electrodes based on empirical observation of the data focuses on the time course of a fixed subset of electrodes, and is thus blind to the evolution of the topography of the sensor space over the time span of the trial. Moreover, the matrix of spatial filters is unique for each subject, thereby factoring

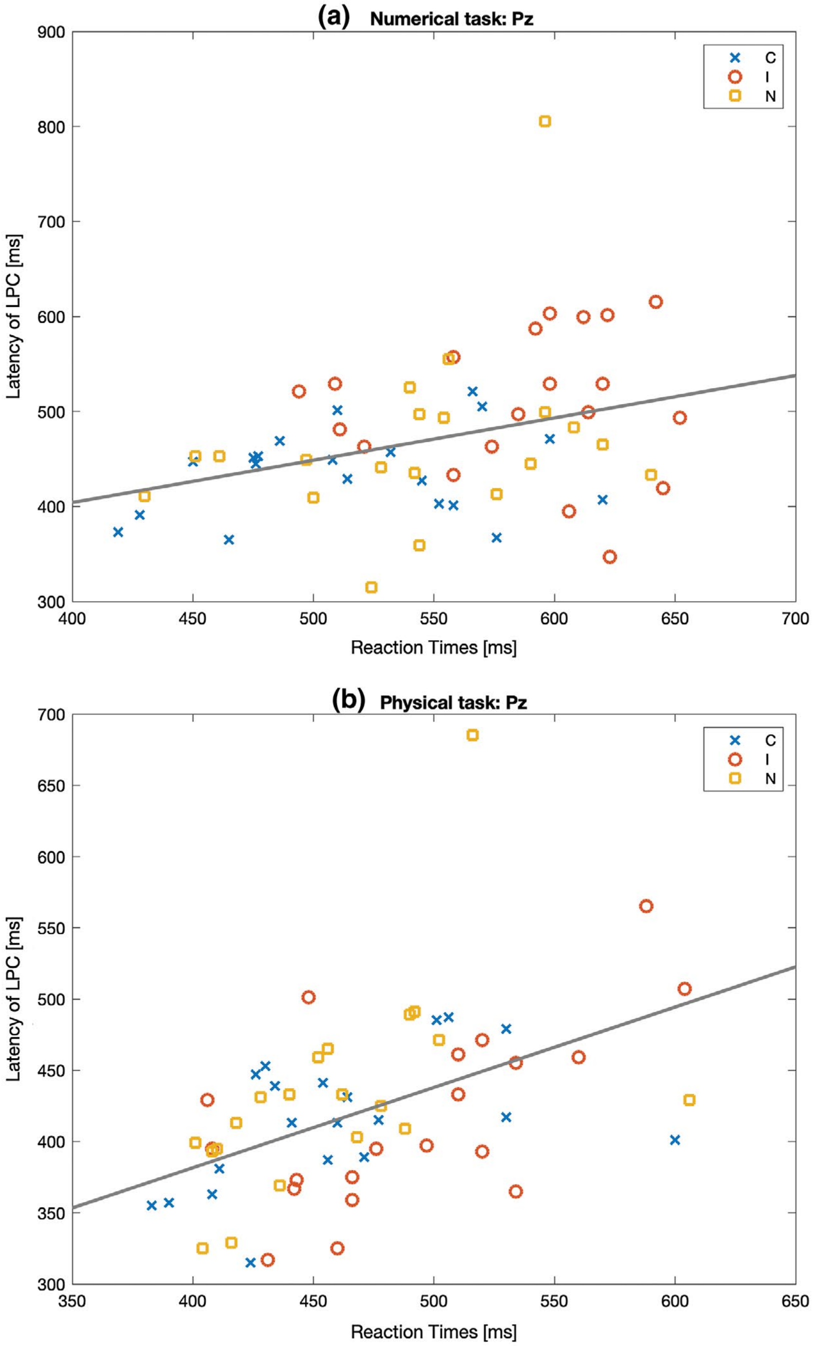

FIGURE 4 Linear regression between LPC peak latency and reaction times of the congruent (C), incongruent (I), and neutral (N) trials for the numerical task (a) and the physical task (b)

out anatomical variability across subjects and focusing specifically on the time course of the experimental effect. EMS filtering attenuates trial-specific noise by projecting the data from each trial onto a matrix of spatial filters derived from all of the other trials. In this sense, EMS filtering is analogous to the sensor noise suppression method (de Cheveigné & Simon, 2008), except that it operates across trials rather than across sensors.

As such, the filter reduced the multichannel EEG recording to a single surrogate trace that preserved the experimental effect. Thirty channels of EEG data (channels × time × trials) were projected into an effect-matched spatial filter producing a set of surrogate signals (time × trials). The idea behind this procedure was to replace a multi-sensor time sample of each trial with one surrogate signal for each trial without prior channel selection. In this way, we ensured that our analysis was not restricted to a region of interest but rather considered all sensor arrays while trying to preserve the experimental variations with a comparison between trials. For algorithm implementation, we followed the procedure of Schurger et al. (2013), except that we conducted *k*-fold stratified cross validation (with five folds) instead of the leave-one-out method proposed in the original article, because the latter option tends to overfit. Before applying the technique, we also normalized the data using z-scores. We compared the mean amplitude of trial surrogates for congruent and incongruent conditions in both tasks with a Wilcoxon

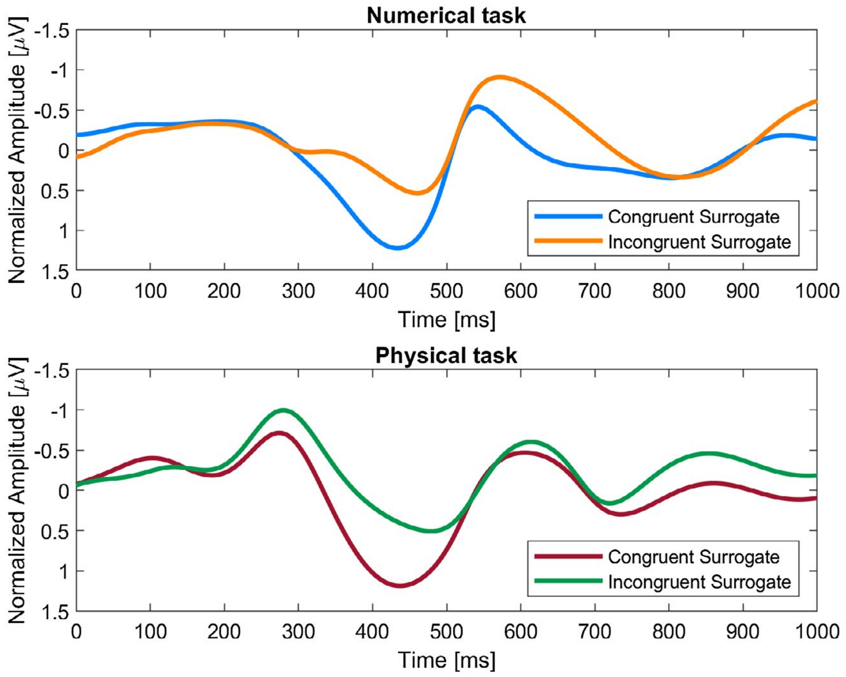

FIGURE 5 Grand average of surrogate trials for the numerical task and the physical task

signed rank test (see Figure 5). For surrogate data, being an artificial manipulation of neurophysiological responses, we preferred a non-parametrical statistical test that did not require any prior assumptions about the data distribution.

We tested the same time windows used in the original time series. When used to test frontal activations, the first time window (200–380 ms) showed a significant congruency effect (congruent vs. incongruent) in both tasks (numerical task: $z = 2.02$, $p < .05$; physical task: $z = 3.57$, $p < .001$). The second time window (380–550 ms) applied on the frontal electrodes maintained this significance (numerical task: $z = 5.13$, $p < .001$; physical task: $z = 4.91$, $p < .001$). The third time window (320–520 ms) used to test parietal effects overlapped with 380–550 ms, and thus returned similar findings (numerical task: $z = 5.05$, $p < .001$; physical task: $z = 5.71$, $p < .001$). For the last time window (520–720 ms), we found a significant congruency effect in the numerical task ($z = 4.06$, $p < .001$) but not in the physical task ($z = 1.09$, $p > .05$) with the surrogate trials. This observation was similar to the mean amplitude tests of ERPs, which reported the same pattern. The analysis of data from the spatially filtered signal could be considered to validate our results for ERP traces, especially in the 520–720 ms time window. In accordance with the results for ERP, the analysis of the surrogate data also showed a later congruency difference for the numerical task, probably due to the difficulty of the experimental task compared with the physical task.

4 | DISCUSSION

In this study, we demonstrated that magnitude in both relevant and irrelevant dimensions affects ERPs and behavioral responses when making judgments. This supports the hypothesis that the processing of physical size and the processing of number overlap in time (Henik & Tzelgov, 1982; MacLeod, 1991). Importantly, including a neutral condition enabled us to better examine how the task-irrelevant feature modulated the processing of the task-relevant attribute. The behavioral results replicated previous findings by showing both facilitation and interference effects in the numerical task but only an interference effect in the physical task (Szűcs & Soltész, 2007). The ERP results in the numerical task showed a facilitation effect on the LPC amplitude. In the physical task, a facilitation effect was identified on the N200 and an interference effect was evident on the N450 and the LPC. The regression results further demonstrated that the LPC latencies were positively correlated with the reaction times for both tasks. Our ERP findings suggest that the task-irrelevant feature influences information processing at multiple levels. Moreover, we also evaluated the time frames of the congruency effects with a spatial filter for both tasks. Given that the congruency effects were consistent between the time-window analysis and the spatially filtered signals, we suggest that this method could be more widely used to validate the experimental effects.

4.1 | The relative processing speed of physical size and number

Our behavioral findings seem consistent with the relative processing speed of physical size and number magnitude (Henik & Tzelgov, 1982), such that the more rapidly processed task-irrelevant size information is able to speed up the more slowly processed numerical information when the two attributes are congruent in the numerical task. In contrast, the more slowly processed task-irrelevant numerical information plays no role in speeding up the processing of size in the physical task, even when the two features are congruent. However, drawing conclusions from this might be premature, as it could stem from differences in the number of possible values per dimension. Most studies (e.g., Kaufmann et al., 2005, 2008; Szűcs & Soltész, 2007), including the current one, have had more possibilities for different numerical values than for different physical sizes in the stimulus pairs. We used eight different digits to create digit pairs, but only three different sizes. Therefore, in consecutive trials, there was a strong chance that one of the physical sizes of the digit pair would be identical to one of the sizes in the previous digit pair. The uncertainty involved in identifying sizes was lower, and hence the judgment of size was easier than that of numerical magnitude, which could explain the faster comparison of physical size. Additionally, the task may have played a role as a top-down mechanism in modulating the relative speed of processing for size and numbers (Dadon & Henik, 2017). For example, as the numerical task encouraged paying attention to magnitude, an attentional bias may have strengthened the (task-relevant) numerical magnitude process (Santens & Verguts, 2011). However, some aspects of information processing such as attention bias may be difficult to capture with discrete measures such as response time or accuracy.

The ERP results from the numerical task suggest that the task-irrelevant size processing was initiated later than the task-relevant number processing, such that the size information played no role in the perceptual matching (indicated by anterior N200) or conflict monitoring (indicated by anterior N450); rather, the size facilitated the stimulus evaluation (indicated by the early LPC). In other words, there was no need to ignore or inhibit the task-irrelevant size information in the earlier processing of numerical magnitude. During stimulus evaluation and response organization, size speeds up the completion of a numerical magnitude judgment when the information on discrete quantity is consistent with that on the abstract quantity of number. Therefore, the LPC amplitude was larger and the latency was shorter. In contrast, longer times are needed to resolve the information conflict generated by task-relevant and task-irrelevant information; thus the LPC amplitude was smaller and the latency was longer.

In the physical task, however, the facilitation effect found on the anterior N200 implies that the task-irrelevant feature—number—had already been processed by this stage. The N200 effect is consistent with literature showing that compatible stimuli (e.g., HHH*H*HHH) elicited a less negative N200 than incompatible stimuli (e.g., SSS*H*SSS) in a flanker task (Eriksen & Eriksen, 1974), which indicates easier perceptual matching for compatible trials. The numerical Stroop paradigm, a two-dimension task, requires the analysis of a conjunction of features. When the discrete quantity and abstract quantity generate compatible activities in mapping to the magnitude representation, the matching process is facilitated. The finding of a N450 interference effect is consistent with the literature on the Stroop test and N450, which has found that incongruent trials elicit a more negative N450 than congruent trials (Lansbergen et al., 2007; Liotti et al., 2000). Studies have suggested that the N450 represents an index of conflict monitoring during the Stroop test (e.g., West & Alain, 1999). The interference effect on the N450 in our study suggests that increased neural resources were used in the detection of conflicting information generated by the task-irrelevant feature. Our findings are in line with the literature indicating that the N450 and N200 reflect different cognitive control processes (for a review, see Tillman & Wiens, 2011). The task-irrelevant feature continually modulates task-relevant processing during the stimulus evaluation and response organization, by showing a smaller and later LPC when the information does not match, and vice versa.

Overall, our results demonstrated that facilitation and interference effects appear at multiple stages of perceptual and response processing in both tasks. This indicates that several different cognitive processes are involved in the phenomena of facilitation and interference, even within a task. The exact processing speeds of the numerical and size dimensions, although largely overlapping, are modulated by the uncertainty of the numerical and size dimensions (Pansky & Algom, 1999, 2002) and/or linguistic properties.

4.2 | Relatively early processing of numerical information

The associations as antagonists to top-down control (AATC) model proposed by Bugg (2014) suggests that top-down control mechanisms are activated only in tasks involving high uncertainty, when it is difficult to use simple associative learning mechanisms. Furthermore, the cognitive control mechanisms can prioritize the processing of the numerical dimension relative to the physical dimension (Dadon & Henik, 2017), although both processes are automatic. This could explain our finding that size information modulated numerical processing during the stimulus evaluation but not perceptual matching in the numerical task, and that numerical magnitude contributed to the size judgment from an earlier stage of information processing. Another possible explanation for numbers being processed earlier in our study might be the linguistic specificity and phonological structure of Chinese: the so-called "Chinese number advantage" (Miura & Okamoto, 2003). In particular, the structure of Chinese number words maps clearly onto the place-value features of the Arabic numeral system, which is consistent with the traditional base-ten numeration system. Research has demonstrated that linguistic properties contribute to differential representation of Arabic numbers in the brain (Nuerk et al., 2005; Pixner et al., 2011) and number processing (Miura & Okamoto, 2003; Tang et al., 2006). Therefore, our results add to the existing literature suggesting that language characteristics may be a factor in the superior numerical processing exhibited by Chinese speakers in numerical Stroop tasks. However, it should also be acknowledged that all the participants in this study have acquired some English; further studies are needed to explore whether the second language experiences or bilingualism would modulate the Arabic number processing and how.

5 | CONCLUSION

Our findings support the hypothesis that multiple factors, such as the specific task requirements, the uncertainty or discriminability of the manipulated dimensions, the proportion of neutral trials (Dadon & Henik, 2017), the proportion of congruent-to-incongruent trials (Borgmann et al., 2011), and even language and/or culture (Cohen Kadosh et al., 2008; Dehaene & Cohen, 2007; Huang et al., 2019; Miura & Okamoto, 2003), all affect the interactions between number and size, and hence the pattern of results in the numerical Stroop task. Ultimately, the human brain is actively adapted to resolve conflict information, and the discussions of early interaction or late interaction between number and size (e.g., Schwarz & Heinze, 1998) might be overly simplified.

Finally, although the ability to perceive and evaluate sizes in discrete quantities might be a primitive system that has been in use throughout evolutionary history (Henik et al., 2012), this does not mean that size is the foundation for number processing (Henik et al., 2017). It remains unclear whether integration of the representation of Arabic numbers with the representation of sizes led to the development of a new magnitude system or whether size representation became accessible for Arabic numbers through evolutionary processes.

ACKNOWLEDGMENT

H. W. Huang and C. M. Huang thank Shih-Ping Huang for his company and indispensable support during the COVID-19 self-quarantine. This work was supported by Academia Sinica Thematic Research Program (AS-103-TP-C04) (for CMH, HWH, and OJLT) and by Taiwan's Ministry of Science and Technology (105-2420-H-009-001-MY2; 106-2410-H-001-024-MY2; 107-2410-H-009-028-MY3; and 108-2321-B-038-005-MY2) (for CMH and OJLT). The work was also supported by the Center for Intelligent Drug Systems and Smart Biodevices (IDS^{2}B) under the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project of the Ministry of Education in Taiwan. This work was also supported by 7200538 (for HWH), and by the Hong Kong Institute for Advanced Study (9360157) (for OJLT and HWH), City University of Hong Kong.

AUTHOR CONTRIBUTION

Hsu-Wen Huang: Conceptualization; Formal analysis; Investigation; Methodology; Project administration; Supervision; Validation; Writing-original draft; Writing-review & editing. **Mauro Nascimben:** Data curation; Formal analysis; Visualization. **Ya-Yi Wang:** Data curation. **Dong-Yang Fong:** Investigation; Resources; Software. **Ovid J.-L. Tzeng:** Conceptualization; Funding acquisition. **Chih-Mao Huang:** Conceptualization; Funding acquisition; Investigation; Methodology; Project administration; Validation.

ORCID

Hsu-Wen Huang https://orcid.org/0000-0001-9677-6855
Chih-Mao Huang https://orcid.org/0000-0002-6209-2575

REFERENCES

Besner, D., & Coltheart, M. (1979). Ideographic and alphabetic processing in skilled reading of English. *Neuropsychologia*, *17*(5), 467–472. https://doi.org/10.1016/0028-3932(79)90053-8

Borgmann, K., Fugelsang, J., Ansari, D., & Besner, D. (2011). Congruency proportion reveals asymmetric processing of irrelevant physical and numerical dimensions in the size congruity paradigm. *Canadian Journal of Experimental Psychology*, *65*(2), 98–104. https://doi.org/10.1037/a0021145

Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society B*, *364*, 1831–1840. https://doi.org/10.1098/rstb.2009.0028

Bugg, J. M. (2014). Conflict-triggered top-down control: Default mode, last resort, or no such thing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 567–587. https://doi.org/10.1037/a0035032

Chang, C. Y., Hsu, S.-H., Pion-Tonachini, L., & Jung, T.-P. (2018). *Evaluation of artifact subspace reconstruction for automatic EEG artifact removal*. Paper presented at the The Institute of Electrical and Electronics Engineers, Inc. (IEEE) Conference, Piscataway.

Cohen Kadosh, R., Cohen Kadosh, K., Linden, D. E., Gevers, W., Berger, A., & Henik, A. (2007). The brain locus of interaction between number and size: A combined functional magnetic resonance imaging and event-related potential study. *Journal of Cognitive Neuroscience*, *19*(6), 957–970. https://doi.org/10.1162/jocn.2007.19.6.957

Cohen Kadosh, R., & Henik, A. (2006). A common representation for semantic and physical properties: A cognitive-anatomical approach. *Experimental Psychology*, *53*(2), 87–94. https://doi.org/10.1027/1618-3169.53.2.87

Cohen Kadosh, R., Henik, A., & Rubinsten, O. (2008). Are Arabic and verbal numbers processed in different ways?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1377–1391. https://doi.org/10.1037/a0013413

Cohen Kadosh, R., Henik, A., Rubinsten, O., Mohr, H., Dori, H., van de Ven, V., Zorzi, M., Hendler, T., Goebel, R., & Linden, D. E. (2005).

Are numbers special? The comparison systems of the human brain investigated by fMRI. *Neuropsychologia*, *43*(9), 1238–1248. https://doi.org/10.1016/j.neuropsychologia.2004.12.017

Cohen Kadosh, R., & Walsh, V. (2009). Numerical representation in the parietal lobes: Abstract or not abstract? *Behavioral and Brain Sciences*, *32*(3–4), 313–328. https://doi.org/10.1017/s0140525x09990938

Coles, M. G. H., & Rugg, M. D. (1995). Event-related brain potentials: An introduction. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition* (pp. 1–26). Oxford University Press.

Dadon, G., & Henik, A. (2017). Adjustment of control in the numerical Stroop task. *Memory & Cognition*, *45*(6), 891–902. https://doi.org/10.3758/s13421-017-0703-6

de Cheveigné, A., & Simon, J. Z. (2008). Sensor noise suppression. *Journal of Neuroscience Methods*, *168*(1), 195–202. https://doi.org/10.1016/j.jneumeth.2007.09.012

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. Oxford University Press.

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, *56*(2), 384–398. https://doi.org/10.1016/j.neuron.2007.10.004

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, *21*(8), 355–361. https://doi.org/10.1016/s0166-2236(98)01263-6

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*, *15*(134), 9–21.

Donchin, E. (1981). Surprise!…Surprise? *Psychophysiology*, *18*(5), 493–513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Fabiani, M., Gratton, G., & Federmeier, K. D. (2007). Event related brain potentials. In J. Cacioppo, L. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 85–119). Cambridge University Press.

Fias, W., Lammertyn, J., Reynvoet, B., Dupont, P., & Orban, G. A. (2003). Parietal representation of symbolic and nonsymbolic magnitude. *Journal of Cognitive Neuroscience*, *15*(1), 47–56. https://doi.org/10.1162/089892903321107819

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, *45*(1), 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x

Henik, A., Gliksman, Y., Kallai, A., & Leibovich, T. (2017). Size perception and the foundation of numerical processing. *Current Directions in Psychological Science*, *26*(1), 45–51. https://doi.org/10.1177/0963721416671323

Henik, A., Leibovich, T., Naparstek, S., Diesendruck, L., & Rubinsten, O. (2012). Quantities, amounts, and the numerical core system. *Frontiers in Human Neuroscience*, *5*(186). https://doi.org/10.3389/fnhum.2011.00186

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition*, *10*(4), 389–395. https://doi.org/10.3758/BF03202431

Huang, C.-M., Doole, R., Wu, C., Huang, H.-W., & Chao, Y.-P. (2019). Culture-related and individual differences in regional brain volumes: A cross-cultural voxel-based morphometry study. *Frontiers in Human Neuroscience*, *13*(9), https://doi.org/10.3389/fnhum.2019.00313

Huang, C.-M., Polk, T. A., Goh, J. O., & Park, D. C. (2012). Both left and right posterior parietal activations contribute to compensatory processes in normal aging. *Neuropsychologia*, *50*(1), 55–66. https://doi.org/10.1016/j.neuropsychologia.2011.10.022

Kaufmann, L., Ischebeck, A., Weiss, E., Koppelstaetter, F., Siedentopf, C., Vogel, S. E., Gotwald, T., Marksteiner, J., & Wood, G. (2008). An fMRI study of the numerical Stroop task in individuals with and without minimal cognitive impairment. *Cortex*, *44*(9), 1248–1255. https://doi.org/10.1016/j.cortex.2007.11.009

Kaufmann, L., Koppelstaetter, F., Delazer, M., Siedentopf, C., Rhomberg, P., Golaszewski, S., Felber, S., & Ischebeck, A. (2005). Neural correlates of distance and congruity effects in a numerical Stroop task: An event-related fMRI study. *NeuroImage*, *25*, 888–898. https://doi.org/10.1016/j.neuroimage.2004.12.041

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*(3), 557–577. https://doi.org/10.1017/S0048577201990559

Lansbergen, M. M., van Hell, E., & Kenemans, J. L. (2007). Impulsivity and conflict in the Stroop task: An ERP study. *Journal of Psychophysiology*, *21*(1), 33–50. https://doi.org/10.1027/0269-8803.21.1.33

Liotti, M., Woldorff, M. G., Perez, R., & Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia*, *38*(5), 701–711. https://doi.org/10.1016/S0028-3932(99)00106-2

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203. https://doi.org/10.1037/0033-2909.109.2.163

Miura, I. T., & Okamoto, Y. (2003). Language supports for mathematics understanding and performance. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 229–242). Lawrence Erlbaum Associates Publishers.

Nuerk, H. C., Weger, U., & Willmes, K. (2005). Language effects in magnitude comparison: Small, but not irrelevant. *Brain and Language*, *92*(3), 262–277. https://doi.org/10.1016/j.bandl.2004.06.107

Pansky, A., & Algom, D. (1999). Stroop and Garner effects in comparative judgment of numerals: The role of attention. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 39–58. https://doi.org/10.1037/0096-1523.25.1.39

Pansky, A., & Algom, D. (2002). Comparative judgment of numerosity and numerical magnitude: Attention preempts automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(2), 259–274. https://doi.org/10.1037/0278-7393.28.2.259

Pinel, P., Dehaene, S., Rivière, D., & LeBihan, D. (2001). Modulation of parietal activation by semantic distance in a number comparison task. *NeuroImage*, *14*(5), 1013–1026. https://doi.org/10.1006/nimg.2001.0913

Pinel, P., Piazza, M., Le Bihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron*, *41*(6), 983–993. https://doi.org/10.1016/S0896-6273(04)00107-2

Pixner, S., Zuber, J., Heřmanová, V., Kaufmann, L., Nuerk, H. C., & Moeller, K. (2011). One language, two number-word systems and many problems: Numerical cognition in the Czech language. *Research in Developmental Disabilities*, *32*(6), 2683–2689. https://doi.org/10.1016/j.ridd.2011.06.004

Polich, J. (2007). Updating p300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Pontifex, M. B., Miskovic, V., & Laszlo, S. (2017). Evaluating the efficacy of fully automated approaches for the selection of eyeblink ICA components. *Psychophysiology*, *54*(5), 780–791. https://doi.org/10.1111/psyp.12827

Posner, M. I., & Snyder, C. R. R. (2004). Attention and cognitive control. In D. A. Balota & E. J. Marsh (Eds.), *Key readings in cognition. Cognitive psychology: Key readings* (pp. 205–223). Psychology Press.

Santens, S., & Verguts, T. (2011). The size congruity effect: Is bigger always more? *Cognition*, *118*(1), 97–113. https://doi.org/10.1016/j.cognition.2010.10.014

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66. https://doi.org/10.1037/0033-295X.84.1.1

Schurger, A., Marti, S., & Dehaene, S. (2013). Reducing multi-sensor data to a single time course that reveals experimental effects. *BMC Neuroscience*, *14*, 122. https://doi.org/10.1186/1471-2202-14-122

Schwarz, W., & Heinze, H.-J. (1998). On the interaction of numerical and size information in digit comparison: A behavioral and event-related potential study. *Neuropsychologia*, *36*(11), 1167–1179. https://doi.org/10.1016/S0028-3932(98)00001-3

Schwarz, W., & Ischebeck, A. (2003). On the relative speed account of number-size interference in comparative judgments of numerals. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(3), 507–522. https://doi.org/10.1037/0096-1523.29.3.507

Szűcs, D., & Soltész, F. (2007). Event-related potentials dissociate facilitation and interference effects in the numerical Stroop paradigm. *Neuropsychologia*, *45*(14), 3190–3202. https://doi.org/10.1016/j.neuropsychologia.2007.06.013

Szűcs, D., & Soltész, F. (2008). The interaction of task-relevant and task-irrelevant stimulus features in the number/size congruency paradigm: An ERP study. *Brain Research*, *1190*, 143–158. https://doi.org/10.1016/j.brainres.2007.11.010

Tang, Y., Zhang, W., Chen, K., Feng, S., Ji, Y., Shen, J., Reiman, E. M., & Liu, Y. (2006). Arithmetic processing in the brain shaped by cultures. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(28), 10775–10780. https://doi.org/10.1073/pnas.0604416103

Tillman, C. M., & Wiens, S. (2011). Behavioral and ERP indices of response conflict in Stroop and flanker tasks. *Psychophysiology*, *48*(10), 1405–1411. https://doi.org/10.1111/j.1469-8986.2011.01203.x

Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 166–179. https://doi.org/10.1037/0278-7393.18.1.166

Verleger, R., Jaśkowski, P., & Wascher, E. (2005). Evidence for an integrative role of P3b in linking reaction to perception. *Journal of Psychophysiology*, *19*(3), 165–181. https://doi.org/10.1027/0269-8803.19.3.165

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*(11), 483–488. https://doi.org/10.1016/j.tics.2003.09.002

West, R., & Alain, C. (1999). Event-related neural activity associated with the Stroop task. *Cognitive Brain Research*, *8*(2), 157–164. https://doi.org/10.1016/S0926-6410(99)00017-8

How to cite this article: Huang H-W, Nascimben M, Wang Y-Y, Fong D-Y, Tzeng OJ-L, Huang C-M. Which digit is larger? Brain responses to number and size interactions in a numerical Stroop task. *Psychophysiology*. 2021;58:e13744. https://doi.org/10.1111/psyp.13744