The role of experience for abstract concepts: Expertise modulates the electrophysiological correlates of mathematical word processing

Laura Bechtold[a], Christian Bellebaum[a], Sophie Egan[a], Marco Tettamanti[b], and Marta Ghio[a]

[a]Heinrich Heine University Düsseldorf, Germany
[b]IRCCS San Raffaele Scientific Institute, Milano, Italy

Author Note

Laura Bechtold, Christian Bellebaum, Sophie Egan, and Marta Ghio, Institute for Experimental Psychology, Heinrich Heine University, Düsseldorf. Marco Tettamanti, Division of Neuroscience, IRCCS San Raffaele Scientific Institute, Milano. Correspondence concerning this article should be addressed to Laura Bechtold, Heinrich Heine University Düsseldorf, Institute for Experimental Psychology, Building 23.03., Universitätsstr. 1, D-40225 Germany, E-Mail: Laura.Bechtold@hhu.de.

Abstract

Embodied theories assign experience a crucial role in shaping conceptual representations. Supporting evidence comes mostly from studies on concrete concepts, where e.g., motor expertise facilitated action concept processing. This study examined experience-dependent effects on abstract concept processing. We asked participants with high and low mathematical expertise to perform a lexical decision task on mathematical and nonmathematical abstract words, while acquiring event-related potentials. Analyses revealed an interaction of expertise and word type on the amplitude of a fronto-central N400 and a centro-parietal late positive component (LPC). For mathematical words, we found a trend for a lower N400 and a significantly higher LPC amplitude in experts compared to non-experts. No differences between groups were found for nonmathematical words. The results suggest that expertise affects the processing stages of semantic integration and memory retrieval specifically for expertise-related concepts. This study supports the generalization of experience-dependent conceptual processing mechanisms to the abstract domain.

Keywords: mathematical words; abstract concepts; semantic memory; expertise; lexical decision; N400; LPC.

1. Introduction

In semantic memory, information derived from our individual experience is stored in form of conceptual representations, which make this knowledge available for cognition, language and action. Theories on the neural underpinnings of semantic memory assign different roles to experience in the acquisition and processing of concepts. Theoretical approaches range from amodal/symbolic to grounded and embodied accounts (for reviews Barsalou, 2008; Meteyard, Cuadrado, Bahrami, & Vigliocco, 2012). The former postulate that initial experiential information is translated into modality-independent representations. In contrast, strongly embodied theories postulate that conceptual processing reactivates experiential information grounded in modality-specific areas, which were activated during the experience with the concepts' referents (Gallese & Lakoff, 2005; Glenberg, 1997). Theories assuming a weaker form of embodiment additionally include higher order convergence zones mediating such reactivation (Galetzka, 2017; Kiefer & Pulvermuller, 2012; Patterson, Nestor, & Rogers, 2007).

A growing body of research provides evidence for an involvement of experiential information from sensory and motor modalities in the representation of concrete concepts, which also reflects their belonging to a specific category (e.g., animals, tools, actions; Binder & Desai, 2011; Ralph, Jefferies, Patterson, & Rogers, 2017). It is, however, not clear whether the idea of grounding can be applied to abstract concepts (e.g., *justice, algebra, to think*), as they refer to entities that we cannot directly experience through our senses (Binder & Desai, 2011; Ralph et al., 2017). According to longstanding theories on semantic concreteness (e.g., *dual coding theory,* Paivio, 1986; *context availability model*, Schwanenflugel & Shoben, 1983), abstract concepts rely exclusively on linguistic information (for a recent review, see Hoffman, 2016). Recent advances within the grounded and embodied cognition framework emphasize the role of experiential aspects referring to social, introspective, affective and magnitude information for abstract concepts (Desai, Reilly, & van Dam, 2018; Ghio, Vaghi, & Tettamanti, 2013; Hoffman, 2016; Troche, Crutch, & Reilly, 2014; Wilson-Mendenhall, Simmons, Martin, & Barsalou, 2013).

Empirical evidence for the contribution of experiential information to abstract concept representation, however, is scarce. This shortage reflects the difficulty in devising an experimental paradigm that addresses individual experience for abstract concepts. Moreover, previous studies rarely used the category-specific approach applied in the research on concrete concepts to examine fine-grained abstract categories (e.g., social, mathematics, mental states; for an example, see Ghio, Vaghi, Perani, & Tettamanti, 2016). One experimental approach to examine the role of experience for concrete concepts has been, indeed, to compare semantic processing in experts versus nonexperts with respect to specific action categories (e.g., Beilock, Lyons, Mattarella-Micke, Nusbaum, & Small, 2008; Locatelli, Gatti, & Tettamanti, 2012). The studies applying this method suggest that expertise, which can be defined as greater proficiency derived from experience or training, leads to an augmented recruitment of experiential modality-specific brain areas as resources for conceptual processing, and affects behavioral responses to verbal stimuli referring to the area of expertise.

The approach of comparing the processing of concepts of a specific category in experts versus nonexperts can be extended to the abstract domain. For this purpose, mathematical concepts (e.g., *multiplication*) seem particularly suitable. These can be considered as a specific abstract category, as suggested by the results of a previous psycholinguistic rating study (Ghio et al., 2013). Within the embodied framework, the hypothesis has been put forward that mathematical concepts are grounded in the same brain areas that were activated during mathematical experience such as calculation and number processing (Wilson-Mendenhall et al., 2013). Accordingly, Wilson-Mendenhall et al. (2013) showed that processing the word *arithmetic* (an abstract mathematical concept) compared to *convince* (an abstract social concept) induced greater activations in brain areas that were also activated during a numerical localizer task, including the intraparietal sulcus and the prefrontal cortex. These areas have been repeatedly shown to underlie mathematical cognition in studies on calculation and number perception (Dehaene, Molko, Cohen, & Wilson, 2004; Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999). The functional network underlying mathematical processing further includes a region in the bilateral inferior temporal cortex, which is more activated by the processing of visually presented Arabic numbers than by the processing of other symbols (i.e., letters or pictures) and has been labeled the visual number form area (Hermes et al., 2017; Shum et al., 2013).

By generalizing from the studies on expertise-induced modulations of action concepts to the mathematical abstract domain, one could hypothesize that the involvement of this mathematics-related prefrontal-intraparietal brain network in processing mathematical concepts would be modulated by individual experience. Evidence for a refinement of parietal areas involved in magnitude-processing by mathematical experience stems from a very recent study on 10 to 12 year old children (Suarez-Pellicioni & Booth, 2018). Another functional magnetic resonance imaging study examined the processing of advanced mathematical statements (e.g., *A finite left-invariant measure over a compact group is bi-invariant*) in professional mathematicians versus nonmathematicians (Amalric & Dehaene, 2016). In mathematicians, performing semantic judgments on these statements specifically induced activations in prefrontal and intraparietal brain regions involved in number processing and calculation. The study by

Amalric and Dehaene (2016) also revealed that activations in the mathematics-related brain regions increased soon after mathematical statement offset and lasted for 15 s.

Previous studies applied event-related potentials (ERPs) to examine more fine-grained temporal dynamics of conceptual category processing (Kiefer, 2001; Lau, Phillips, & Poeppel, 2008). A largely used electrophysiological indicator of semantic processing is the N400 component. It has been associated not only with context-dependent semantic anomalies (Kutas & Hillyard, 1983), but also with the processing of isolated words or pictures (Lau et al., 2008). The N400 has been shown to be sensitive to semantic categories, as it differed between visual and auditory-related concepts (Bastiaansen, Oostenveld, Jensen, & Hagoort, 2008), between natural objects and artifacts (Kiefer, 2001), as well as between concrete and abstract concepts (Adorni & Proverbio, 2012; Barber, Otten, Kousta, & Vigliocco, 2013; Holcomb, Kounios, Anderson, & West, 1999; Kounios & Holcomb, 1994). These ERP effects have been interpreted in terms of category-specific access to lexical and semantic information, and are thus thought to reflect how conceptual knowledge is represented and which type of experiential information is being retrieved (Kiefer, 2001; but see Hauk, 2016). Furthermore, the N400 effect has been shown for incongruous vs. congruous arithmetic problems, which suggests that the processing of arithmetic and semantic anomalies relies on at least partly overlapping mechanisms (Niedeggen & Rösler, 1999; Niedeggen, Rösler, & Jost, 1999; for a positive component involved in arithmetic processing in this time window see, e.g., Dehaene, 1996).

Another ERP component that has been found to be sensitive to conceptual category differences, especially for the distinction between abstract and concrete concepts (e.g., Adorni & Proverbio, 2012; Kanske & Kotz, 2007), is a late positive component (LPC). A fronto-central LPC has been interpreted as in terms of either mental imagery (Kanske & Kotz, 2007) or topdown control of semantic memory (Adorni & Proverbio, 2012). A more centro-parietal LPC has been suggested to reflect the recollection of individual experience (Strozak, Bird, Corby, Frishkoff, & Curran, 2016) and the retrieval of arithmetic facts involved in solving complex but not simple mathematical problems (Kiefer & Dehaene, 1997). In addition, a recent study demonstrated that the amplitude of the LPC was affected by arithmetic anomalies (Dickson & Federmeier, 2017; 400 ms to 600 ms).

In the present study, we specifically aimed to provide evidence for an experience-dependent modulation of mathematical concept processing with respect to mathematical expertise, which we objectively evaluated by administering a math test. We focused on the temporal dynamics of this modulation by measuring ERPs of participants with high versus low mathematical expertise performing a lexical decision task. To avoid a lack of effective comprehension of complex mathematical statements in nonexperts, we used single words instead of sentences (see Amalric & Dehaene, 2016). In a pre-experimental rating with nonexperts, these mathematical words' familiarity ratings did not differ significantly from those of nonmathematical abstract words, which served as a standard of comparison in our ERP study. This design allowed us to test the specificity of mathematical expertise in modulating the processing of words referring to mathematical abstract concepts. We hypothesized that, if mathematical expertise contributes to shaping conceptual representations of mathematical words, its modulatory effect on their conceptual processing might already become apparent in the N400 and in the LPC.

2. Method

2.1 Participants

All 46 participants of the present study were students, between 18 and 30 years old, had normal or corrected-to-normal vision, no history of psychiatric or neurological diseases and were right-handed. One participant had to be excluded due to technical problems during the data acquisition. Two additional participants were excluded from the statistical analysis because their mean LPC amplitudes deviated by more than three standard deviations from the mean of their respective group at three electrode sites (see Section 2.4.2.2 for a detailed description of the ERP analysis and scoring). The participants were recruited from different disciplines (mathematics, natural sciences, economics, psychology and humanities) at Heinrich Heine University Düsseldorf, in order to have a heterogeneous sample with respect to the scope of the mathematical education. Each participant completed a math test to quantify his/her mathematical expertise (see Section 2.2.1 for details). Participants with a test score of at least 7 points (total: 12 points) were assigned to the group with high mathematical expertise (HiEx). The HiEx group consisted of 23 participants (14 males, mean age = 22.8 years, *SD* = 3.3). Participants with test scores below 7 points were assigned to the group with low mathematical expertise (LoEx). This group consisted of 20 participants (10 males, mean age = 22.8 years, *SD* = 3.0). In the math test, participants of the HiEx group reached a significantly higher mean score (*M* = 8.4 points, *SD* = 1.0) than the participants of the LoEx group, (*M* = 3.3 points, *SD* = 2.3), as revealed by an independent samples *t*-test, $t(24.816) = 9.305$, $p < .001$, $d = 2.991$.

The study is in line with the declaration of Helsinki and was approved by the ethics committee of the Faculty of Mathematics and Natural Sciences at Heinrich Heine University. All participants gave their written informed consent prior to their participation, for which they received monetary compensation or course credit.

2.2 Material

2.2.1 Assessment of mathematical expertise

A math test assessed the level of mathematical expertise. It contained 12 mathematical problems (four arithmetical, four algebraic and four analytical problems). No time limit was set for completing the test. Two independent raters evaluated the participants' performance on the test. For each of the 12 problems one point was given for the correct solution; half a point if the approach to the problem was correct but the result was incorrect. The first 14 participants (2 HiEx, 12 LoEx) underwent the test after the EEG acquisition. In order to obtain a comparable number of experts and nonexperts, however, we subsequently targeted recruitment towards students of mathematics and natural sciences, and administered the test before the EEG acquisition. The following 29 participants (21 HiEx and 8 LoEx) were tested with this modified order of the procedure. In this second phase of the data acquisition, four volunteers did not undergo the EEG experiment because they did not reach the required score for the HiEx group.

2.2.2 Stimuli

For the lexical decision task (see Section 2.3), we used 31 mathematical (MAT) words, 31 nonmathematical (NONMAT) abstract words and 62 pseudo-words (see Table S1 in the supplementary material for a complete list). The MAT words included mathematical terms (e.g., *multiplication* or *mathematics*), but not number words. The NONMAT words mostly referred to mental or emotional states (e.g., *thought* or *fear*). We matched the words for length (number of letters; MAT: *M* = 8.42, *SD* = 2.20; NONMAT: *M* = 7.74, *SD* = 2.00; *t*(60) = 1.268, *p* = .210, *d* = 0.323) and lexical frequency (as assessed via the Wortschatz Lexikon of the University of Leipzig, http://wortschatz.uni-leipzig.de; MAT: *M* = 14601.06, *SD* = 76282.49; NONMAT: *M* = 9766.16, *SD* = 17304.96; *t*(60) = 0.344, *p* = .732, *d* = 0.087). Importantly, we matched MAT and NONMAT words for the psycholinguistic variables concreteness, abstractness, valence, and familiarity based on a rating by an independent sample of 64 German-speaking participants. MAT and NONMAT words differed significantly only in ratings of arousal (see Table 1, left).

Table 1

Pre- and Post-experimental rating of psycholinguistic variables

Scale	Word type	Pre-experimental rating				Follow-up rating	
		independent raters	*df* ^{a}	*t*	*p*	LoEx	HiEx
Concreteness	MAT	3.51 (0.72)	60	0.939	.351	3.23 (1.83)	2.59 (1.57)
	NONMAT	3.32 (0.88)				2.94 (1.32)	2.24 (0.89)
Abstractness	MAT	5.30 (0.65	53.130	.881	.382	5.08 (1.79)	3.97 (1.90)
	NONMAT	5.12 (0.95)				4.55 (1.38)	4.58 (1.59)
Valence ^{b}	MAT	3.94 (0.29)	32.101	-0.033	.974	3.92 (0.40)	4.08 (0.79)
	NONMAT	3.95 (1.58)				4.10 (0.25)	4.00 (0.63)
Arousal	MAT	2.06 (0.44)	37.961	-9.934	<.001	1.45 (0.48)	2.35 (1.63)
	NONMAT	4.32 (1.19)				3.57 (1.19)	3.39 (1.24)
Familiarity	MAT	4.97 (0.61)	60	-1.140	.259	4.58 (1.87)	5.71 (1.20)
	NONMAT	5.16 (0.70)				5.68 (1.14)	5.04 (1.64)

Note. Means (*SD*) and inferential statistics for the independent samples *t*-tests of the pre-experimental stimulus validation rating are presented on the left side. The right side shows the respective follow-up rating results for participants with low (LoEx, *n* = 13) and high (HiEx, *n* = 14) mathematical expertise. The ratings were performed on 1-7 Likert scales for concreteness, abstractness, valence, arousal and familiarity, for the mathematical (MAT) and nonmathematical (NONMAT) words.

[a] Degrees of freedom were corrected in case of unequal variances.

[b] Valence was rated on a -3 (negative) to +3 (positive) scale, with 0 (neutral). For better comparability, values were transformed to a 1 (negative) to 7 (positive) scale with 4 depicting neutral values.

To create word-like pseudo-words (e.g., *Hatrip*), we used the pseudo-word generator *Wuggy* (Keuleers & Brysbaert, 2010) with the German language module. The 31 MAT and 31 NONMAT words served as input, from which the program generated one pseudo-word each. The generation parameters restricted the output pseudo-words to match the input words in length of subsyllabic segments, letter length, transition frequencies between letters, and two out of three subsyllabic segments.

2.3 Experimental procedure

2.3.1 Lexical decision task

We applied a lexical decision task, which is a rather implicit task, in order to prevent overt attention to the semantic category manipulation. This task should therefore induce brain activations that reflect aspects of knowledge that are intrinsic to the representation of the concepts. When applied with word-like pseudo-words, as it was done in the present study (see Section 2.2.2), the lexical decision task has been shown to successfully induce semantic processing (Barber et al., 2013; Binder et al., 2003). For each participant, the acquisition took place in a dimly lit, electrically shielded EEG laboratory. Each trial began with a fixation cross that remained for a random interval of 1200 ms to 1600 ms, followed by a (pseudo-) word presented on the screen for 800 ms. Then, a blank screen was shown with a duration between 300 ms and 500 ms, followed by a screen prompting the participants' response. The participants' task was to distinguish between words and pseudo-words by pressing a button at the end of each trial. The response buttons (left and right) were randomly assigned to the decision options (word and pseudo-word) between trials. This procedure aimed to avoid motor artifacts caused by preparatory finger movements. If no response was given within 10 seconds, the next trial started automatically. The inter-trial interval had a duration of 500 ms, throughout which a blank screen was shown. All stimuli were presented on a black background in a white sans-serif font (Arial) of the size 20 pt.

Participants were instructed to look at the fixation cross and to try to avoid any movement. They first completed six practice trials with three words and three pseudo-words not included in the experiment. All 31 MAT, 31 NONMAT and 62 pseudo-words were presented twice in two separate experimental runs, adding up to 124 trials per run and to a total number of 248 trials. The order of the presentation of the words and pseudo-words was randomized within each run. During each experimental run, participants had the opportunity to take self-paced breaks after every 16 trials. The software Presentation (version 17.0, Neurobehavioral Systems Inc., Albany, CA, USA) was used for stimulus presentation and response recording. We used a Windows 10 Dell Intel Premium PC, a 22” LED Dell monitor with 1680*1050 pixel resolution and a refresh rate of 60 Hz. Responses were given via two response buttons (left/right) on the Cedrus RB-844 response pad (Cedrus Corporation, San Pedro, California).

2.3.2 EEG recording

Twenty-eight Ag/AgCl ring electrodes were used to record electrical potentials on the scalp. They were positioned on a BrainCap textile softcap (Brainproducts GmbH, Germany) following the extended 10-20 system (Chatrian, Lettich, & Nelson, 1985; electrode sites were F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T7, C3, Cz, C4, T8, CP3, CPz, CP4, P7, P3, Pz, P4, P8, PO7, PO3, POz, PO4 and PO8). The ground electrode was attached to site AFz, the linked reference electrodes to the mastoids. Careful scalp preparation kept impedances below 5 kΩ. Four additional electrodes recorded eye movements: one above and one below the left eye, as well as two at the outer canthi of the eyes. The EEG data was recorded with a BrainAmp DC amplifier (Brainproducts GmbH, Germany), a sampling rate of 1000 Hz, a lowpass filter of 1000 Hz and no highpass filter on a Windows 10 Dell Intel Premium PC with the Brain Vision Recorder software (version 1.20.0506, Brain Products GmbH, Germany).

2.4 Data Analysis

Statistical analysis was conducted with IBM SPSS statistics (version 23.0, IBM Corporation, USA). For all inferential statistics, an alpha level of .05 was assumed. Degrees of freedom were adjusted according to the Greenhouse-Geisser and Welch-Satterthwaite methods, in the case of violations of sphericity and homogeneity, respectively. Follow-up tests for significant interactions as well as multiple correlations were corrected for the false discovery rate (FDR) with the procedure introduced by Benjamini and Hochberg (1995). Pseudo-words were not considered in the analyses, as we were not interested in lexicality effects. As measures of effect size we report η_p^2 or Cohen's *d* (calculated with JASP, version 0.8.3.1, JASP Team(2018)), where appropriate.

2.4.1 Behavioral data

Accuracy in the lexical decision task was calculated as the percentage of correct responses of all given responses. To analyze accuracy, we applied a 2x2 mixed ANOVA with the between-subjects factor Group (HiEx, LoEx) and the within-subjects factor Word Type (MAT, NONMAT).

2.4.2 EEG data

2.4.2.1 Data preprocessing

Data preprocessing was conducted with the Brain Vision Analyzer software (version 2.1, Brainproducts GmbH, Germany). We applied a Butterworth zero phase filter with a low cutoff of 0.1 Hz (time constant: 1.59, slope of 24 dB/Oct) and a high cutoff of 30 Hz, both with a slope of 48 dB/Oct. Additionally, a notch filter for the frequency of 50 Hz was applied to eliminate power supply hum. Then, a fast independent component analysis (ICA) with classical sphering on a 120 s excerpt of the data of each participant was used to discard one or two components related to blink artifacts. The continuous EEG was then segmented into epochs from 300 ms before to 1200 ms after onset of the presented words. After a baseline correction that subtracted the mean signal of the 200 ms interval prior to stimulus onset from the data, an automatic procedure detected artifacts of non-cerebral origin at the 15 electrodes used in the statistical analyses (see Section 2.4.2.2 ERP data analysis). The parameters were the following: The maximal allowed voltage step from one data point to the next was 50 µV, the minimal/maximal allowed difference of amplitude values between the highest and the lowest data point within 100 ms intervals was 0.1 µV and 100 µV, respectively, and the minimally/maximally allowed amplitudes were ± 100 µV. Next, all artifact-free trials were averaged for each participant, separately for the two conditions of MAT and NONMAT words. On average, 59.8 MAT (*SD* = 2.9) and 59.6 NONMAT word trials (*SD* = 4.0) were used for the averaged ERPs.

2.4.2.2 ERP data analysis

Visual inspection of the ERP waveforms, averaged across participants, revealed a frontally pronounced N400, in line with the literature on N400 concreteness effects (Adorni &

Proverbio, 2012; Barber et al., 2013; Holcomb et al., 1999; Kounios & Holcomb, 1994; Strozak et al., 2016), while the LPC was more positive over posterior electrodes (compare, e.g., Kandhadai & Federmeier, 2010a; Strozak et al., 2016). The N400 was quantified as the mean amplitude in the time window between 350 ms and 450 ms for each of the nine electrodes of a fronto-central cluster (F3, Fz, F4, FC3, FCz, FC4, C3, Cz, C4). The LPC component, which was quantified as the mean amplitude between 500 ms and 700 ms after stimulus onset, was analyzed for a centro-parietal cluster of nine electrodes (C3, Cz, C4, CP3, CPz, CP4, P3, Pz, P4). Notably, a fronto-central P2 preceded the N400 (see Figure 1A), and seemed to have a slightly higher amplitude in the LoEx group. In order to examine this potential group difference and its potential impact on the subsequent N400 and LPC results, we also extracted the P2 peak amplitude, which was defined as the local maximum between 170 ms and 300 ms at the nine fronto-central electrode sites. The P2, N400 and LPC were then analyzed in separate 2x2x3x3 mixed ANOVAs with the between-subjects factor Group (HiEx, LoEx) and the within-subject factors Word Type (MAT, NONMAT), Frontality (frontal, fronto-central, central for the N400 and P2; central, centro-parietal, parietal for the LPC) and Laterality (left, midline, right). Effects of the topographical factors Frontality and Laterality are reported only if they interacted significantly with at least one of the non-topographical factors.

3. Results

3.1 Behavioral Data

Across participants, the mean accuracy in the lexical decision task was very high in all groups and conditions (at least 97.8%). Statistical analysis revealed that the Group did not have a significant effect on accuracy, $F(1, 41) = 4.595$, $p = .194$, $\eta_p^2 = .041$. The effect of the Word

Type was significant, $F(1, 41) = 5.476$, $p = .024$, $\eta_p^2 = .118$, with a lower accuracy for MAT ($M = 98.5\%$, $SD = 1.8\%$) than for NONMAT words ($M = 99.1\%$, $SD = 1.3\%$). The Group x Word Type interaction was also significant, $F(1, 41) = 9.574$, $p = .004$, $\eta_p^2 = .189$. Dependent samples *t*-tests revealed that the HiEx group had a similar accuracy for MAT ($M = 99.2\%$, $SD = 1.2\%$) and NONMAT words ($M = 98.9\%$, $SD = 1.3\%$), $t(22) = 0.536$, $p = .598$, $d = 0.112$, while for the LoEx group accuracy was significantly lower for MAT ($M = 97.8\%$, $SD = 2.0\%$) than NONMAT words ($M = 99.4\%$, $SD = 1.3\%$), $t(19) = -3.866$, $p = .002$, $d = -0.864$.

3.2 ERP Data

Figure 1 depicts the ERPs elicited by MAT and NONMAT words, separately for the two groups, at all electrode sites involved in the analyses, as well as pooled across the nine electrodes used for the N400 and P2 analyses (Figure 1 A), and the nine electrodes used for the LPC analysis (Figure 1 B).

insert Figure 1 here

3.2.1 P2

In the HiEx group the mean P2 peak amplitude was 6.251 µV ($SD = 3.108$ µV) for MAT and 6.419 µV ($SD = 3.241$ µV) for NONMAT words. The LoEx group showed mean amplitudes of 6.513 µV ($SD = 2.416$ µV) for MAT and 6.703 µV ($SD = 3.170$ µV) for NONMAT words. Neither Group, $F(1, 41) = 0.098$, $p = .756$, $\eta_p^2 = .002$, nor Word Type, $F(1, 41) = 0.467$, $p = .498$, $\eta_p^2 = .011$, nor the Group x Word Type interaction, $F(1, 41) = 0.003$, $p = .954$, $\eta_p^2 < .001$, had a significant effect on P2 amplitudes. None of the interactions of the factors Group and Word Type with the topographical factors were significant (all $p \geq .064$). We thus assumed that the results reported in the following were not affected by the P2 component.

3.2.2 N400

Neither Group, $F(1, 41) = 2.070$, $p = .158$, $\eta_p^2 = .048$, nor Word Type, $F(1, 41) = 0.092$, $p = .763$, $\eta_p^2 = .002$, had a significant main effect on the N400 amplitudes. Notably, the Group x Word Type interaction was significant, $F(1, 41) = 6.993$, $p = .012$, $\eta_p^2 = .146$ (for descriptive statistics see the bar graph in Figure 1 A, right). The descriptive pattern showed a cross-over interaction with a reduced N400 for MAT compared to NONMAT words in the HiEx group (*mean difference* = 0.745 µV, *SD* = 1.724 µV) and an enhanced (less positive) N400 for MAT compared to NONMAT words in the LoEx group (*mean difference* = -0.592 µV, *SD* = 1.568 µV). To examine this significant interaction further, we first applied dependent samples *t*-tests to compare the two word types within each group. However, the N400 amplitude difference failed to reach significance in the HiEx group, $t(22) = 2.073$, $p = .100$, $d = 0.432$, as well as in the LoEx group, $t(19) = -1.688$, $p = .108$, $d = -0.377$. Focusing on between-group differences, independent samples *t*-tests revealed a trend towards reduced N400 amplitudes in response to MAT words for the HiEx compared to the LoEx group (*mean difference* = 2.060 µV, *SD* = 3.127 µV), $t(41) = 2.154$, $p = .074$, $d = 0.659$. The groups clearly did not differ regarding the N400 for NONMAT words (*mean difference* = 0.723 µV, *SD* = 3.405 µV), $t(41) = 0.694$, $p = .491$, $d = 0.212$. The three-way interaction Group x Frontality x Laterality was significant, $F(2.914, 119.464) = 3.047$, $p = .033$, $\eta_p^2 = .069$. Independent samples *t*-tests, comparing the two groups at each electrode site, revealed that the differences in N400 amplitudes were largest, albeit not significant, at electrode sites C4 (*mean difference*: 2.246 µV, *SE* = 0.975 µV), $t(41) = 2.303$, $p = .234$, $d =$

0.704, and FC4 (*mean difference*: 2.083 µV, *SE* = 0.944 µV), $t(41) = 2.207$, $p = .149$, $d = 0.675$ (all other $p \geq .374$). Descriptively, amplitudes were lower (more positive) in the HiEx group. No other interactions with the factors Group and Word Type and the topographical factors were significant (all $p \geq .055$).

3.2.3 LPC

The LPC was significantly affected by Group (more positive amplitudes in the HiEx group, $F(1, 41) = 5.419$, $p = .025$, $\eta_p^2 = .117$), as well as by Word Type (more positive amplitudes for MAT words, $F(1, 41) = 6.678$, $p = .013$, $\eta_p^2 = .140$). The Group x Word Type interaction was significant as well, $F(1, 41) = 4.972$, $p = .031$, $\eta_p^2 = .108$ (for descriptive statistics see the bar graph in Figure 1 B, right). Dependent samples *t*-tests revealed that the HiEx group had a significantly more positive LPC amplitude when processing MAT words compared to NONMAT words (*mean difference* = 1.477 µV, *SD* = 2.192 µV), $t(22) = 3.231$, $p = .008$, $d = 0.674$. In the LoEx group, MAT and NONMAT words did not elicit significantly different LPC amplitudes (*mean difference* = 0.109 µV, *SD* = 1.769 µV), $t(19) = 0.275$, $p = .786$, $d = 0.061$. Additional independent samples *t*-tests revealed that the HiEx group showed a significantly higher LPC amplitude than the LoEx group in response to MAT words (*mean difference* = 2.836 µV, *SD* = 2.948 µV), $t(41) = 3.147$, $p = .006$, $d = 0.962$. LPC amplitudes in response to NONMAT words did not differ significantly between groups (*mean difference* = 1.467 µV, *SD* = 2.948 µV), $t(41) = 1.409$, $p = .166$, $d = 0.431$. The Word Type x Frontality interaction also reached significance, $F(1.424, 58.375) = 3.865$, $p = .040$, $\eta_p^2 = .086$. In order to explore this interaction, we applied dependent samples *t*-tests comparing the amplitudes elicited by the two word types for each level of Frontality across groups. MAT words elicited significantly higher (more positive) LPC amplitudes than NONMAT words at parietal (*mean difference*: 0.988 µV, *SD* = 2.167 µV) and centro-parietal (*mean difference*: 0.872 µV, *SD* = 2.086 µV) electrode sites, $t(42) = 2.990$, $p = .014$, $d = 0.456$ and $t(42) = 2.742$, $p = .014$, $d = 0.418$, respectively. The comparison was not significant at central electrode sites, $p = .053$. There were no further significant interactions of the factors Word Type and/or Group with the topographical factors (all $p \geq .162$).

3.2.4 Correlation of ERP data with the mathematical test score

To explore the relationship between expertise and ERP indicators of conceptual processing further, we correlated the participants' math test score with the MAT-NONMAT amplitude difference of the N400 and LPC (pooled over the nine electrodes that entered the analysis for each component) by means of two-sided Pearson correlations. N400 ($r = .487$, $p = .002$), as well as LPC ($r = .442$, $p = .003$) amplitude differences significantly (FDR corrected) correlated with the math test scores.

3.3 Follow-up psycholinguistic rating

To verify whether the degree of mathematical expertise of participants in the HiEx and LoEx group was also reflected by the psycholinguistic evaluation of the MAT and NONMAT words, we collected ratings of the experimental stimuli from the participants in our EEG study in a follow-up online rating. This rating included the same 7-point Likert-scales for concreteness, abstractness, valence, familiarity, and arousal as in the pre-experimental rating performed by a separate sample of participants (see Section 2.2.2). Fourteen participants from the HiEx and 13 from the LoEx group participated in the follow-up rating. Importantly, also in this sub-sample the performance in the math test differed significantly between the HiEx (*M* = 8.3, *SD* = 0.7) and LoEx group (*M* = 3.5, *SD* = 2.5), *t*(13.755) = 6.716, $p < .001$, *d* = 2.673.

3.3.1 Follow-up rating results

Descriptive statistics of the rating results are displayed in Table 1 (right). For each scale, ratings were analyzed by applying a 2 (Group: HiEx, LoEx) x 2 (Word Type: MAT, NONMAT) mixed ANOVA. Consistent with the pre-experimental rating, we did not find any significant main or interaction effects for the concreteness and valence scores (all $p > .160$).

There were no main effects of Group or Word Type on the abstractness and familiarity ratings (all $p \geq .368$). However, we found a significant Group x Word Type interaction for abstractness, *F*(1, 25) = 4.484, *p* = .044, $\eta_p^2 = .152$, and familiarity, *F*(1, 25) = 13.144, *p* = .001, $\eta_p^2 = .345$. Concerning abstractness ratings, the interaction was likely due to the fact that the pattern was descriptively reversed between the two groups. However, dependent samples *t*-tests did neither reveal a significant difference between MAT and NONMAT words in the LoEx group (*mean difference* = 0.531, *SD* = 1.472), *t*(12) = 1.301, *p* = .218, *d* = 0.361, nor in the HiEx group (*mean difference* = -0.615, *SD* = 1.341), *t*(13) = -1.716, *p* = .218, *d* = -0.459. Focusing on differences between the two groups, independent samples *t*-tests comparing the MAT words (*mean difference* = -1.112, *SD* = 1.846) and NONMAT words (*mean difference* = 0.035, *SD* = 1.490) did not reveal any significant differences either, *t*(25) = -1.563, *p* = .262, *d* = -0.602 and *t*(25) = 0.060, *p* = .953, *d* = 0.023, respectively. Concerning the familiarity ratings, dependent samples *t*-tests showed that the LoEx group rated MAT words lower than NONMAT words (*mean difference* = -1.099, *SD* = 1.170), *t*(12) = -3.386, *p* = .010, *d* = -0.939, while in the HiEx group MAT words yielded descriptively higher scores than NONMAT words, although this difference did not reach significance, $t(13) = 1.848$, $p = .087$, $d = 0.494$. Independent samples *t*-tests revealed no significant differences between the two groups for MAT (*mean difference* = 1.124, *SD* = 1.586), $t(20.224) = 1.841$. $p = .160$, $d = 0.721$ and NONMAT words (*mean difference* = -0.636, *SD* = 1.399), $t(23.213) = -1.164$, $p = .250$, $d = -0.448$.

Concerning the arousal ratings, we replicated the main effect of Word Type observed in the pre-experimental rating, $F(1, 25) =36.402$, $p < .001$, $\eta_p^2 = .593$, with MAT words receiving a significantly lower mean arousal rating than NONMAT words. Neither the main effect of Group, nor the Group x Word Type interaction was significant, $F(1, 25) = 0.864$, $p = .362$, $\eta_p^2 = .033$ and $F(1, 25) = 4.196$, $p = .051$, $\eta_p^2 = .144$, respectively.

3.3.2 Correlations between follow-up rating and ERP data

The results of the follow-up rating suggested that the degree of mathematical expertise was reflected by abstractness and familiarity ratings of MAT versus NONMAT words. For this reason, we examined whether abstractness and familiarity ratings correlated with the modulations of the ERPs that we observed. Specifically, we performed two-sided Pearson correlations (FDR corrected) for the MAT-NONMAT word rating differences (separately for abstractness and familiarity) with the MAT-NONMAT ERP amplitude differences (separately for N400 and LPC, pooled over nine electrodes). Note that only the sub-sample of the 14 HiEx and 13 LoEx participants who completed the follow-up ratings could be considered in this analysis. The results of the correlation analyses revealed that abstractness and familiarity rating differences neither correlated significantly with the N400 ($r = -.036$, $p = .903$ and $r = .050$, $p = .903$, respectively) nor the LPC amplitude difference ($r = .212$, $p = .466$ and $r = .314$, $p = .466$, respectively).

4. Discussion

The current study aimed to extend previous evidence for experience-dependent neural representations of concrete concepts (Kiefer & Pulvermuller, 2012) to abstract concepts, by testing whether the individual degree of mathematical expertise (high versus low) specifically modulates the linguistic processing of mathematical concepts. Consistent with our hypotheses, we found a significant interaction of the factors Group and Word Type on the amplitudes of a fronto-central N400 and a centro-parietal LPC. For the N400 component the resolution of the interaction revealed that the processing of MAT words led to a trend-level reduction of the N400 amplitude in participants of the HiEx group compared to the LoEx group, while processing nonmathematical words clearly did not differ between groups. Concerning the LPC component, a significantly more pronounced LPC was found for the processing of mathematical words in the HiEx group compared to the LoEx group, again with no differences between groups for the processing of nonmathematical words. This pattern of results indicates that the degree of expertise with mathematical concepts influenced semantic processing differentially over time.

Single word studies suggest that the N400 amplitude is sensitive to the ease of lexical access and activation of semantic information from long-term memory, and thus reflects aspects of semantic categorization (Kutas & Federmeier, 2000). Specifically, a reduction of the N400 amplitude is considered to reflect either a facilitated activation of semantic features associated with the lexical item, or a reduced need to integrate information from multiple semantic regions (Lau et al., 2008). A higher N400 in response to arithmetic incongruences has been interpreted to reflect a higher processing effort comparable to semantically anomalous sentences (Niedeggen & Rösler, 1999; Niedeggen et al., 1999). Thus, the trend for relatively reduced N400 amplitudes elicited by MAT word processing in the HiEx compared to the LoEx group might be interpreted in terms of a relatively reduced processing effort for MAT words in participants with a high level of mathematical experience. Importantly, however, the N400 amplitude has been sensitive to multiple factors that modulate lexical access (Kutas & Federmeier, 2000; Lau et al., 2008). This raises the question which of the possible factors led to the relatively reduced N400 amplitudes elicited by the processing of mathematical compared to nonmathematical words in the HiEx group. The psycholinguistic rating scores collected in this study can help to exclude some potentially confounding variables, as MAT and NONMAT words were matched for concreteness and valence. Arousal ratings were significantly lower for MAT words in the pre-experimental as well as in the follow-up rating. However, as the arousal ratings for the two word types did not differ between HiEx and LoEx participants, and as arousal has been found to have an impact on word processing only in interaction with valence (Bayer, Sommer, & Schacht, 2010; Yao et al., 2016), an influence of arousal on our ERP results seems unlikely.

Another potentially confounding factor is word familiarity: Studies have shown that less familiar words result in higher N400 amplitudes (Bader & Mecklinger, 2017; Barber, Vergara, & Carreiras, 2004; Lau et al., 2008; Rugg, 1990; Strozak et al., 2016; Vergara-Martinez, Comesana, & Perea, 2017; Vergara-Martinez & Swaab, 2012). Although MAT and NONMAT words were counterbalanced for their frequency of occurrence and familiarity, as measured in a pre-experimental validation rating with an independent sample of nonexperts, a follow-up rating showed that MAT words were indeed rated as less familiar than NONMAT words by LoEx participants, while there was no significant difference to the HiEx participants' familiarity ratings. However, we found that familiarity ratings did not correlate with the N400 amplitude. It can therefore be ruled out that the N400 amplitude modulation by expertise solely reflected differences in word familiarity.

Considering our experimental manipulation, another factor potentially influencing the N400 is the extent of mathematical experience of the participants, which has probably enriched the knowledge they associate with mathematical concepts. We quantified participants' mathematical expertise in terms of their math test performance and showed that it indeed correlated with the N400 amplitude. The HiEx participants' higher mean test score could thus serve as a complementary measure of familiarity with mathematical concepts, providing a more objective, content-based criterion than the merely subjective amount of exposure assessed via familiarity ratings. The test scores might reflect qualitatively different experiences with mathematical concepts, including the successful application of solution strategies. Thus, it seems likely that the HiEx group was more familiar not with the MAT words per se but with the underlying MAT concepts, which reflects the core of their expertise.

To our knowledge, there have not been any ERP studies investigating the role of expertise in conceptual processing of abstract concepts so far. However, indirect evidence that the processing of abstract concepts associated with an experientially enriched content might modulate the N400 amplitude comes from studies on abstract emotional concepts. Stronger experience-dependent emotional content of abstract words facilitated their processing, as reflected in faster reaction times (Kanske & Kotz, 2007; Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011), and reduced N400 amplitudes (Kanske & Kotz, 2007; Trauer, Kotz, & Muller, 2015). Accordingly, the consolidated mathematical experience of the HiEx participants could have enriched their mathematical conceptual representations, leading to the relatively reduced fronto-central N400 amplitude. The current study, however, cannot disentangle whether this effect reflects a facilitated lexical access and feature retrieval, or rather world knowledge integration (Lau et al., 2008). These aspects could be addressed in future research, e.g., by systematically varying a given sentential context for the mathematical words.

Yet another interpretation for the N400 differences between the HiEx and LoEx groups is that the extent of mathematical experience might affect what type(s) of semantic features are associated with mathematical concepts. Category-specific N400 modulation effects have been interpreted as indicative of differences in the type of experience-dependent semantic information (e.g., visual, action) activated by concrete concepts (Adorni & Proverbio, 2009; Kellenbach, Wijers, & Mulder, 2000; Kiefer, 2001, 2005). In studies comparing concrete and abstract concepts, higher N400 amplitudes at frontal electrode sites have been interpreted as indicating stronger sensorimotor integration processes for concepts with a more pronounced inherent multimodality (Adorni & Proverbio, 2012; Barber et al., 2013; Holcomb et al., 1999; Kounios & Holcomb, 1994). In line with these previous studies, we can speculate that the fronto-central N400 modulations we observed for the processing of mathematical concepts reflect a stronger integration of multimodal (i.e., visuospatial and sensorimotor) information in LoEx than HiEx participants. So far, there is only limited evidence for the contribution of multimodal information to the representation of mathematical concepts (Ghio et al., 2013).

However, previous research found abstract number concepts (e.g., *nine*) to be grounded in visuospatial (Spatial Numerical Association of Response Codes [SNARC] effect; Dehaene, Bossini, & Giraux, 1993; Fischer, 2008; Marghetis, Landy, & Goldstone, 2016) and in sensorimotor brain areas, as derived from either spatial number mapping or finger counting habits (Domahs, Moeller, Huber, Willmes, & Nuerk, 2010). Notably, Cipora et al. (2016) found that the SNARC effect was absent in participants with mathematical expertise, which suggests a reduced involvement of multimodal information in abstract numerical representations for mathematical experts (Cipora et al., 2016; but see Sella, Sader, Lolliot, & Cohen Kadosh, 2016). Similarly, the relatively smaller N400 amplitude for MAT words in the HiEx group of this study might be interpreted in terms of such a reduced involvement of multimodal information in mathematical conceptual processing. MAT words received descriptively higher abstractness ratings from LoEx than HiEx participants in the follow-up rating of this study, which seems contradictory at first. However, this might again reflect the actual mathematical experience with the MAT concepts, which made the words seem less abstract to HiEx participants (see also above). Future research might use more fine-grained ratings of abstractness or even a feature production task in order to explore the content that participants with different levels of expertise assign to abstract mathematical concepts.

For the HiEx participants, in turn, mathematical concepts might rely more on mathematics-related semantic information, and therefore activate a brain network specialized on mathematical processing (Wilson-Mendenhall et al., 2013). Intriguingly, an ERP study showed that, when compared to other concrete categories, numerals were the only category that did not elicit a negativity but rather a bilateral parietal positivity (Dehaene, 1995). Such a parietal positivity was also found in number magnitude comparison (Dehaene, 1996) and multiplication (Kiefer & Dehaene, 1997). These results are also consistent with recent findings of the recruitment of number processing and calculation brain areas for the processing of mathematical statements in mathematicians (Amalric & Dehaene, 2016). Although we also observed a descriptively reduced N400 amplitude associated with the processing of mathematical concepts in the HiEx group, a bilateral parietal positivity did not become apparent in that time interval in our data, but in the later one of the LPC.

This later parietal positivity was more pronounced for MAT words in the HiEx group. In addition to this interaction, we also found significant main effects of Group and Word Type. However, as pairwise comparisons resolving the interaction showed that the main effects were driven by the higher LPC amplitudes in response to MAT words in the HiEx group, we will not interpret them separately. The parietal pronunciation of the LPC in the present study might suggest that the mathematical network identified by Amalric and Dehaene (2016), in which parietal structures play a prominent role, was recruited during mathematical conceptual processing in the HiEx group. Kiefer and Dehaene (1997) reported a longer lasting and bilateral, instead of only left-hemispheric, positivity over parietal areas for complex versus simpler mathematical problems. The authors interpreted this problem size effect as reflecting the retrieval of mathematical knowledge from parietal areas. A recent study with children also found a stronger bilateral parietal activation for a harder (but not easier) magnitude processing task and interpreted it to reflect a refined representation of quantity induced by mathematical experience (Suarez-Pellicioni & Booth, 2018). Such a recall of mathematical knowledge from parietal areas might also have caused the LPC modulation observed in the current study. Our LPC results might thus reflect the stronger reactivation or integration of mathematical knowledge in mathematical experts. This interpretation should be considered with caution, however, given that the spatial information provided by the scalp topography is limited. Furthermore, although there is some evidence that the LPC is sensitive to mathematical stimulus processing, the direction of this modulation in our study is not consistent with an interpretation in terms of ease of retrieval of semantic features (Dickson & Federmeier, 2017; Guthormsen et al., 2016).

Alternatively, the parietal LPC modulation might be interpreted in terms of recollection of individual experience associated with the conceptual content. While a more fronto-central

LPC for the processing of concrete versus abstract conceptual categories has been interpreted as indicating mental imagery (Kanske & Kotz, 2007), higher parietal LPC amplitudes have more consistently been linked to strategic, conscious memory processes. This interpretation is based on studies with healthy subjects (Kandhadai & Federmeier, 2010a; Strozak et al., 2016), as well as aphasic (Swaab, Brown, & Hagoort, 1998) and amnestic (Olichney et al., 2000) patients. Usually, LPC modulations depend on tasks explicitly demanding memory recollection (Fischer-Baum, Dickson, & Federmeier, 2014; Kandhadai & Federmeier, 2010b), while the lexical decision task we applied in this study is a rather implicit task. This suggests that rather than being task-related, LPC modulations in our study might be related to the degree of individual experience, with mathematical expertise motivating the recollection of information related to the mathematical concepts. The higher LPC amplitudes elicited by MAT words in the HiEx group might therefore result from explicit, strategic memory retrieval, driven by a stronger recollection of experiential information (Daltrozzo, Wioland, & Kotchoubey, 2007; Guthormsen et al., 2016; Kandhadai & Federmeier, 2010a) or recollection-based reanalysis (Van Petten & Luka, 2012). This recollection of consolidated experiential information might also be required for mental simulations involved in higher level conceptual processing (Barsalou, 2008).

In conclusion, the present study provides evidence for experience-dependent modulations of mathematical concept processing, reflected by specific modulations of mathematical word processing. The relatively reduced N400 amplitudes elicited by mathematical words in the expert group could be the result of a less effortful conceptual processing as well as a reduced reliance on multimodal integration. The more positive LPC elicited by mathematical words in the experts possibly reflects an enhanced retrieval of experiential information in their area of expertise.

Taken together, our results speak for a contribution of mathematical experience to shaping and processing mathematical concepts.

Acknowledgments

We want to thank Catherine Ries, Anna Solowiej and Duygu Tarasi for their help with the data acquisition.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Figure Caption

Figure 1. *Group and Word Type effects on N400 and LPC amplitudes*.

The central part shows the grand average ERPs elicited by mathematical (MAT) and nonmathematical (NONMAT) word processing in the high (HiEx, $n = 23$) and low (LoEx, $n = 20$) expertise group at the electrode sites included, respectively, in the N400 and LPC analyses. A. Left: ERPs pooled over the nine fronto-central electrodes. Shaded area marks the N400 time window (350-450 ms). Right: Mean amplitudes of the N400 separately for the levels of Group and Word Type. B. Left: ERPs pooled over the nine centro-parietal electrodes. Shaded area marks the LPC time window (500-700 ms). Right: Mean amplitudes of the LPC separately for the levels of Group and Word Type. Error bars represent ± one standard error.

References

Adorni, R., & Proverbio, A. M. (2009). New insights into name category-related effects: is the Age of Acquisition a possible factor? *Behav Brain Funct,* 5, 33.

Adorni, R., & Proverbio, A. M. (2012). The neural manifestation of the word concreteness effect: an electrical neuroimaging study. *Neuropsychologia,* 50(5), 880-891.

Amalric, M., & Dehaene, S. (2016). Origins of the brain networks for advanced mathematics in expert mathematicians. *Proc Natl Acad Sci U S A,* 113(18), 4909-4917.

Bader, R., & Mecklinger, A. (2017). Separating Event-related Potential Effects for Conceptual Fluency and Episodic Familiarity. *J Cogn Neurosci,* 29(8), 1402-1414.

Barber, H. A., Otten, L. J., Kousta, S. T., & Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain Lang,* 125(1), 47-53.

Barber, H. A., Vergara, M., & Carreiras, M. (2004). Syllable-frequency effects in visual word recognition: evidence from ERPs. *Neuroreport,* 15(3), 545-548.

Barsalou, L. W. (2008). Grounded cognition. *Annu Rev Psychol,* 59, 617-645.

Bastiaansen, M. C., Oostenveld, R., Jensen, O., & Hagoort, P. (2008). I see what you mean: theta power increases are involved in the retrieval of lexical semantic information. *Brain Lang,* 106(1), 15-28.

Bayer, M., Sommer, W., & Schacht, A. (2010). Reading emotional words within sentences: the impact of arousal and valence on event-related potentials. *Int J Psychophysiol,* 78(3), 299-307.

Beilock, S. L., Lyons, I. M., Mattarella-Micke, A., Nusbaum, H. C., & Small, S. L. (2008). Sports experience changes the neural processing of action language. *Proc Natl Acad Sci U S A,* 105(36), 13269-13273.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological),* 57(1), 289-300.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends Cogn Sci,* 15(11), 527-536.

Binder, J. R., McKiernan, K. A., Parsons, M. E., Westbury, C. F., Possing, E. T., Kaufman, J. N., et al. (2003). Neural correlates of lexical access during visual word recognition. *J Cogn Neurosci,* 15(3), 372-393.

Chatrian, G. E., Lettich, E., & Nelson, P. L. (1985). Ten Percent Electrode System for Topographic Studies of Spontaneous and Evoked EEG Activities. *American Journal of EEG Technology,* 25(2), 83-92.

Cipora, K., Hohol, M., Nuerk, H. C., Willmes, K., Brozek, B., Kucharzyk, B., et al. (2016). Professional mathematicians differ from controls in their spatial-numerical associations. *Psychol Res,* 80(4), 710-726.

Daltrozzo, J., Wioland, N., & Kotchoubey, B. (2007). Sex differences in two event-related potentials components related to semantic priming. *Arch Sex Behav,* 36(4), 555-568.

Dehaene, S. (1995). Electrophysiological evidence for category-specific word processing in the normal human brain. *Neuroreport,* 6(16), 2153-2157.

Dehaene, S. (1996). The organization of brain activations in number comparison: event-related potentials and the additive-factors method. *J Cogn Neurosci,* 8(1), 47-68.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The Mental Representation of Parity and Number Magnitude. *Journal of Experimental Psychology: General,* 122(3), 371-396.

Dehaene, S., Molko, N., Cohen, L., & Wilson, A. J. (2004). Arithmetic and the brain. *Curr Opin Neurobiol,* 14(2), 218-224.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science,* 284(5416), 970-974.

Desai, R. H., Reilly, M., & van Dam, W. (2018). The multifaceted abstract brain. *Philos Trans R Soc Lond B Biol Sci,* 373(1752).

Dickson, D. S., & Federmeier, K. D. (2017). The language of arithmetic across the hemispheres: An event-related potential investigation. *Brain Res,* 1662, 46-56.

Domahs, F., Moeller, K., Huber, S., Willmes, K., & Nuerk, H. C. (2010). Embodied numerosity: implicit hand-based representations influence symbolic number processing across cultures. *Cognition,* 116(2), 251-266.

Fischer-Baum, S., Dickson, D. S., & Federmeier, K. D. (2014). Frequency and regularity effects in reading are task dependent: Evidence from ERPs. *Lang Cogn Neurosci,* 29(10), 1342-1355.

Fischer, M. H. (2008). Finger counting habits modulate spatial-numerical associations. *Cortex,* 44(4), 386-392.

Galetzka, C. (2017). The Story So Far: How Embodied Cognition Advances Our Understanding of Meaning-Making. *Front Psychol,* 8, 1315.

Gallese, V., & Lakoff, G. (2005). The Brain's concepts: the role of the Sensory-motor system in conceptual knowledge. *Cogn Neuropsychol,* 22(3), 455-479.

Ghio, M., Vaghi, M. M., & Tettamanti, M. (2013). Fine-grained semantic categorization across the abstract and concrete domains. *PLoS One,* 8(6), e67090.

Ghio, M., Vaghi, M. M. S., Perani, D., & Tettamanti, M. (2016). Decoding the neural representation of fine-grained conceptual categories. *Neuroimage,* 132, 93-103.

Glenberg, A. M. (1997). What memory is for. *Behav Brain Sci,* 20(1), 1-19; discussion 19-55.

Guthormsen, A. M., Fisher, K. J., Bassok, M., Osterhout, L., DeWolf, M., & Holyoak, K. J. (2016). Conceptual Integration of Arithmetic Operations With Real-World Knowledge: Evidence From Event-Related Potentials. *Cogn Sci,* 40(3), 723-757.

Hermes, D., Rangarajan, V., Foster, B. L., King, J. R., Kasikci, I., Miller, K. J., et al. (2017). Electrophysiological Responses in the Ventral Temporal Cortex During Reading of Numerals and Calculation. *Cereb Cortex,* 27(1), 567-575.

Hoffman, P. (2016). The meaning of 'life' and other abstract words: Insights from neuropsychology. *J Neuropsychol,* 10(2), 317-343.

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: an electrophysiological investigation. *J Exp Psychol Learn Mem Cogn,* 25(3), 721-742.

Kandhadai, P., & Federmeier, K. D. (2010a). Automatic and controlled aspects of lexical associative processing in the two cerebral hemispheres. *Psychophysiology,* 47(4), 774-785.

Kandhadai, P., & Federmeier, K. D. (2010b). Hemispheric differences in the recruitment of semantic processing mechanisms. *Neuropsychologia,* 48(13), 3772-3781.

Kanske, P., & Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Res,* 1148, 138-148.

Kellenbach, M. L., Wijers, A. A., & Mulder, G. (2000). Visual semantic features are activated during the processing of concrete words: event-related potential evidence for perceptual semantic priming. *Brain Res Cogn Brain Res,* 10(1-2), 67-75.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: a multilingual pseudoword generator. *Behav Res Methods,* 42(3), 627-633.

Kiefer, M. (2001). Perceptual and semantic sources of category-specific effects: event-related potentials during picture and word categorization. *Mem Cognit,* 29(1), 100-116.

Kiefer, M. (2005). Repetition-priming modulates category-related effects on event-related potentials: further evidence for multiple cortical semantic systems. *J Cogn Neurosci,* 17(2), 199-211.

Kiefer, M., & Dehaene, S. (1997). The Time Course of Parietal Activation in Single-digit Multiplication: Evidence from Event-related Potentials. *Mathematical Cognition,* 3(1), 1-30.

Kiefer, M., & Pulvermuller, F. (2012). Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *Cortex,* 48(7), 805-825.

Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *J Exp Psychol Learn Mem Cogn,* 20(4), 804-823.

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: why emotion matters. *J Exp Psychol Gen,* 140(1), 14-34.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn Sci,* 4(12), 463-470.

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Mem Cognit,* 11(5), 539-550.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nat Rev Neurosci,* 9(12), 920-933.

Locatelli, M., Gatti, R., & Tettamanti, M. (2012). Training of manual actions improves language understanding of semantically related action sentences. *Front Psychol,* 3, 547.

Marghetis, T., Landy, D., & Goldstone, R. L. (2016). Mastering algebra retrains the visual system to perceive hierarchical structure in equations. *Cogn Res Princ Implic,* 1(1), 25.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: a review of embodiment and the neuroscience of semantics. *Cortex,* 48(7), 788-804.

Niedeggen, M., & Rösler, F. (1999). N400 Effects Reflect Activation Spread During Retrieval of Arithmetic Facts. *Psychological Science,* 10(3), 271-276.

Niedeggen, M., Rösler, F., & Jost, K. (1999). Processing of incongruous mental calculation problems: evidence for an arithmetic N400 effect. *Psychophysiology,* 36(3), 307-324.

Olichney, J. M., Van Petten, C., Paller, K. A., Salmon, D. P., Iragui, V. J., & Kutas, M. (2000). Word repetition in amnesia. Electrophysiological measures of impaired and spared memory. *Brain,* 123 (Pt 9), 1948-1963.

Paivio, A. (1986). *Mental representations: A dual-coding approach.* Oxford, UK: Oxford University Press.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat Rev Neurosci,* 8(12), 976-987.

Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nat Rev Neurosci,* 18(1), 42-55.

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Mem Cognit,* 18(4), 367-379.

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 9(1), 82-102.

Sella, F., Sader, E., Lolliot, S., & Cohen Kadosh, R. (2016). Basic and advanced numerical performances relate to mathematical expertise but are fully mediated by visuospatial skills. *J Exp Psychol Learn Mem Cogn,* 42(9), 1458-1472.

Shum, J., Hermes, D., Foster, B. L., Dastjerdi, M., Rangarajan, V., Winawer, J., et al. (2013). A brain area for visual numerals. *J Neurosci,* 33(16), 6709-6715.

Strozak, P., Bird, C. W., Corby, K., Frishkoff, G., & Curran, T. (2016). FN400 and LPC memory effects for concrete and abstract words. *Psychophysiology,* 53(11), 1669-1678.

Suarez-Pellicioni, M., & Booth, J. R. (2018). Fluency in symbolic arithmetic refines the approximate number system in parietal cortex. *Hum Brain Mapp*.

Swaab, T. Y., Brown, C., & Hagoort, P. (1998). Understanding ambiguous words in sentence contexts: electrophysiological evidence for delayed contextual selection in Broca's aphasia. *Neuropsychologia,* 36(8), 737-761.

Trauer, S. M., Kotz, S. A., & Muller, M. M. (2015). Emotional words facilitate lexical but not early visual processing. *BMC Neurosci,* 16, 89.

Troche, J., Crutch, S., & Reilly, J. (2014). Clustering, hierarchical organization, and the topography of abstract and concrete nouns. *Front Psychol,* 5, 360.

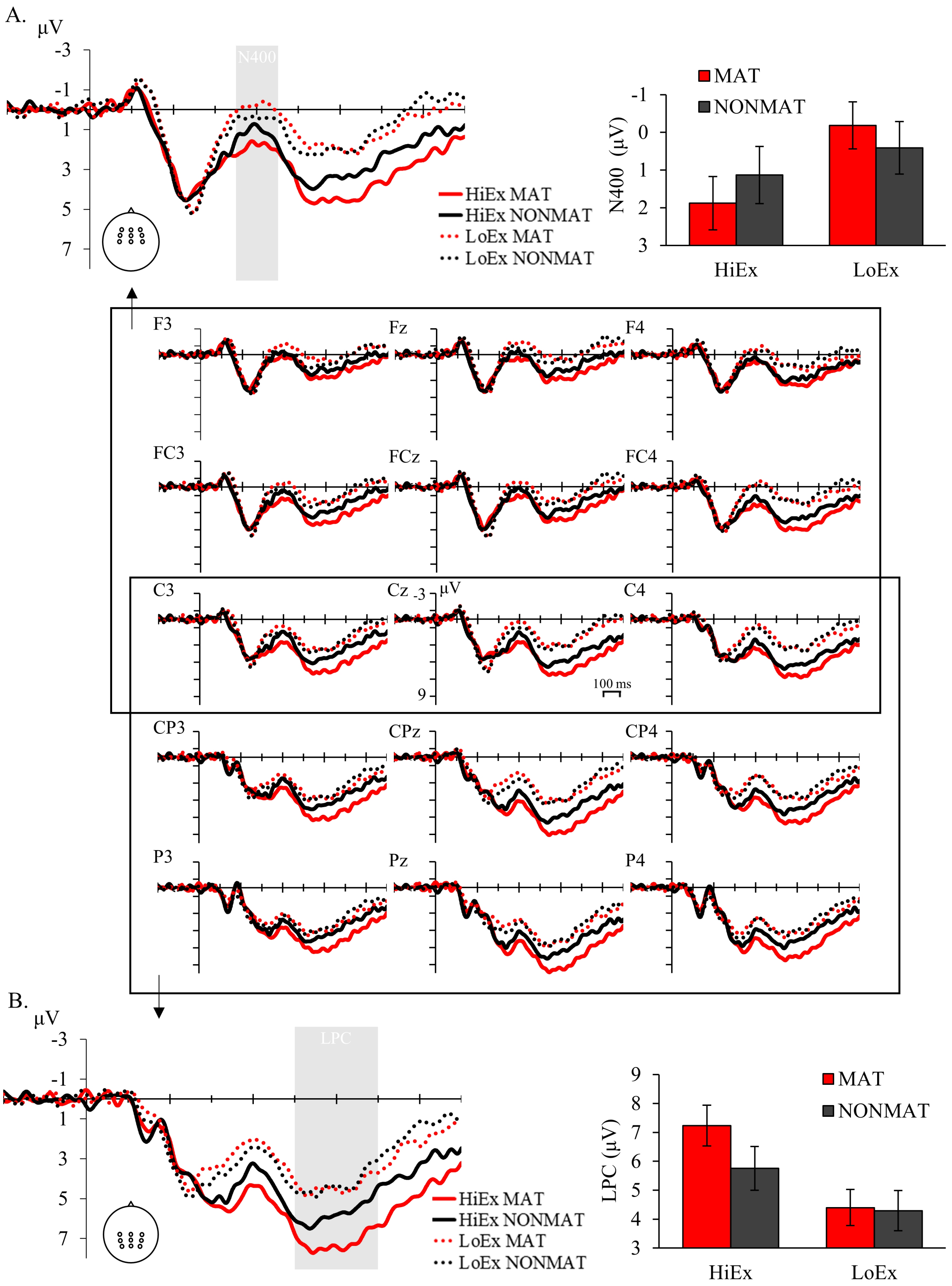
Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: benefits, costs, and ERP components. *Int J Psychophysiol,* 83(2), 176-190.

Vergara-Martinez, M., Comesana, M., & Perea, M. (2017). The ERP signature of the contextual diversity effect in visual word recognition. *Cogn Affect Behav Neurosci,* 17(3), 461-474.

Vergara-Martinez, M., & Swaab, T. Y. (2012). Orthographic neighborhood effects as a function of word frequency: an event-related potential study. *Psychophysiology,* 49(9), 1277-1289.

Wilson-Mendenhall, C. D., Simmons, W. K., Martin, A., & Barsalou, L. W. (2013). Contextual processing of abstract concepts reveals neural representations of nonlinguistic semantic content. *J Cogn Neurosci,* 25(6), 920-935.

Yao, Z., Yu, D., Wang, L., Zhu, X., Guo, J., & Wang, Z. (2016). Effects of valence and arousal on emotional word processing are modulated by concreteness: Behavioral and ERP evidence from a lexical decision task. *Int J Psychophysiol,* 110, 231-242.


A.
μV
-3
-1
1
3
5
7
N400
HiEx MAT
HiEx NONMAT
LoEx MAT
LoEx NONMAT
MAT
NONMAT
N400 (μV)
-1
0
1
2
3
HiEx
LoEx
F3
Fz
F4
FC3
FCz
FC4
C3
Cz
-3
μV
9
C4
100 ms
CP3
CPz
CP4
P3
Pz
P4
B.
μV
-3
-1
1
3
5
7
LPC
HiEx MAT
HiEx NONMAT
LoEx MAT
LoEx NONMAT
9
8
7
6
5
4
3
LPC (μV)
MAT
NONMAT
HiEx
LoEx

Statement of significance

This study investigates the role of experience on processing abstract mathematical concepts. By applying event-related potentials, we demonstrate that the level of mathematical expertise (experts vs. nonexperts) specifically affects automatic and strategic stages of mathematical word processing. These results provide evidence of experience-dependent mechanisms contributing to abstract concept processing.

Supplementary Material

Table S1

Complete list of experimental stimuli

MAT	NONMAT	Pseudo words	
Addition (*addition*)	Absicht (*intention*)	Urmirian	Ungicht
Algebra (*algebra*)	Ankunft (*arrival*)	Augetro	Abspukt
Algorithmus (*algorithm*)	Anliegen (*request*)	Exgarilltum	Wureigen
Analysis (*analysis*)	Effekt (*effect*)	Uharynis	Esfetz
Diagonale (*diagonal*)	Einbildung (*imagination*)	Toabanase	Finkoldung
Division (*division*)	Ewigkeit (*eternity*)	Zühisan	Afigseit
Exponent (*exponent*)	Furcht (*fear*)	Wuvoneug	Fulchz
Geometrie (*geometry*)	Gedächtnis (*memory*)	Neobekrei	Gedäumtbas
Gleichung (*equation*)	Gedanke (*thought*)	Pleispukt	Veginke
Integral (*integral*)	Grund (*reason*)	Wureblol	Vruns
Koeffizient (*coefficient*)	Horror (*horror*)	Toezmiziofs	Korrär
Mathematik (*mathematics*)	Illusion (*illusion*)	Bajehsatif	Exjubiän
Matrix (*matrix*)	Intention (*intention*)	Hatrip	Hudenrian
Maximum (*maximum*)	Kameradschaft (*comradeship*)	Taßigur	Pamerabschohr
Multiplikation (*multiplication*)	Kommunikation (*communication*)	Buntiprebotian	Vespudipation
Nenner (*denominator*)	Leistung (*performance*)	Zenzer	Zeirufst
Normierung (*standardization*)	Loyalität (*loyalty*)	Forseirung	Zygalitat
Ordnungssytem (*classification* system)	Profit (*profit*)	Fultungssößtem	Flofät
Potenz (*power*)	Reflexion (*reflection*)	Poteun	Dejehcian
Primzahl (*prime*)	Schande (*shame*)	Brimzard	Phranbe
Proportion (*proportion*)	Schema (*scheme*)	Stokartian	Streha
Prozent (*percent*)	Scherz (*joke*)	Flozets	Schenz
Quotient (*quotient*)	Schmerz (*pain*)	Blorielz	Schmefs
Statistik (*statistics*)	Trennung (*separation*)	Staristaf	Krentums
Stochastik (*stochastics*)	Tugend (*virtue*)	Stortastif	Imsend
Subtraktion (*subtraction*)	Vermächtnis (*legacy*)	Sunträhtine	Vermüchtbas
Tangente (*tangent*)	Vision (*vision*)	Nabtente	Niriün
Teiler (*divisor*)	Wahnsinn (*madness*)	Leiles	Wallginn
Term (*term*)	Wunsch (*wish*)	Tefs	Fubsch
Vektor (*vector*)	Zustand (*condition*)	Ziktor	Tustind
Winkel (*angle*)	Zweifel (*doubt*)	Mingel	Pleimel

Note. All mathematical (MAT) and nonmathematical (NONMAT) words and the pseudo-words used in the lexical decision task. English translations are provided in parentheses.