Department of Education and Psychology
Freie Universität Berlin

When smaller is more – investigating the interplay between continuous sensory cues and numerical information

Dissertation

submitted for the degree of

Doctor of Natural Sciences
(Dr. rer. nat.)

M.Sc.
Judit Pekár

Berlin, February 2022

Fachbereich Erziehungswissenschaft und Psychologie
der Freien Universität Berlin

When smaller is more – investigating the interplay between continuous sensory cues and numerical information

Dissertation
zur Erlangung des akademischen Grades
Doktorin der Naturwissenschaften
(Dr. rer. nat.)

Vorgelegt von

M.Sc.
Judit Pekár

Berlin, Februar 2022

1st Reviewer/Erstgutachterin:

Prof. Dr. Annette Kinder

2nd Reviewer/Zweitgutachter:

Prof. Dr. Bert Reynvoet

Date of disputation: 09 June 2022

Table of content

Summary

Research on numerical cognition is not limited to symbolic numbers and mathematics but it also includes discrete and continuous magnitudes. Continuous magnitudes are ubiquitous in nature and serve as important cues in everyday life situations. When one tries to choose the plate with more cookies in the cafeteria, they usually do not count the cookies but rather arrive at a fair estimate by comparing such continuous magnitudes. For example, nine cookies on a plate will occupy a larger area and have to be placed denser to each other than five cookies. Recent research has shown that, as opposed to the classical view, the processing of symbolic numbers and non-symbolic numerosities is not independent from such sensory cues. The present dissertation consists of two studies that investigate what psychological processes underlie the interaction between sensory cues and numerical information.

Study 1 aimed to replicate and extend the findings of Gebuis & Reynvoet who systematically manipulated the relationship between continuous and discrete magnitudes in a non-symbolic numerical comparison task. The main goal was to assess the stability and the robustness of the influence of sensory cues on numerical comparisons as the originally reported patterns suggest a complex interaction between these two kinds of information that are difficult to reconcile with the classic views on numerical processing. Indeed, the results confirmed that continuous magnitudes have a complex effect on numerical judgements and that their interaction can be either due to incomplete inhibition or due to integration of continuous magnitudes during numerical tasks.

Study 2 turned to symbolic numbers and investigated whether inhibition underlies the interaction of continuous sensory properties and numerical information. To this end a

novel paradigm was introduced that allowed to investigate well-established electrophysiological correlates of inhibition with numerical stimuli. The results provide evidence that inhibition underlies the interaction between sensory cues and numerical information. Additionally, they show that the paradigm introduced in Study 2 may suitable to investigate these processes across different developmental stages and numeracy levels.

Zusammenfassung

Die Forschung zur numerischen Kognition beschränkt sich nicht nur auf symbolische Zahlen und Mathematik, sondern umfasst auch diskrete Anzahlen und kontinuierliche Mengen. Kontinuierliche Mengen sind in der Natur allgegenwärtig und dienen als wichtige Hinweise in alltäglichen Lebenssituationen. Wenn man versucht, den Teller mit mehr Keksen in der Cafeteria auszuwählen, zählt man normalerweise nicht die Kekse, sondern gelangt zu einer fairen Schätzung, indem man ihre kontinuierliche Ausdehnung vergleicht. Beispielsweise nehmen neun Kekse auf einem Teller eine größere Fläche ein und müssen dichter zueinander platziert werden, als fünf Kekse. Neuere Forschungen haben gezeigt, dass die Verarbeitung symbolischer Zahlen und nicht-symbolischer Zahlen, im Gegensatz zur klassischen Sichtweise, nicht unabhängig von solchen sensorischen Hinweisen ist. Die vorliegende Dissertation besteht aus zwei Studien, die untersuchen, welche psychologischen Prozesse der Wechselwirkung zwischen sensorischen Hinweisen und numerischen Informationen zugrunde liegen.

Studie 1 zielte darauf ab, die Ergebnisse von Gebuis & Reynvoet, wer die Beziehung zwischen kontinuierlichen und diskreten Größen in einer nicht symbolischen numerischen Vergleichsaufgabe systematisch manipulierten, zu replizieren und zu erweitern. Das Hauptziel bestand darin, die Stabilität und Robustheit des Einflusses sensorischer Hinweise auf numerische Vergleiche zu bewerten, da die ursprünglich berichteten Muster eine komplexe Wechselwirkung zwischen diesen beiden Arten von Informationen nahelegen, die sich nur schwer mit den klassischen Ansichten zur numerischen Verarbeitung vereinbaren lassen. In der Tat bestätigten die Ergebnisse, dass kontinuierliche Größen einen komplexen Einfluss auf numerische Beurteilungen haben

und dass ihre Wechselwirkung entweder auf eine unvollständige Inhibition oder auf die Integration kontinuierlicher Größen während numerischer Aufgaben zurückzuführen sein kann.

Studie 2 wandte sich symbolischen Zahlen zu und untersuchte, ob der Interaktion von kontinuierlichen sensorischen Eigenschaften und numerischer Information inhibitorische Prozesse zugrunde liegen. Zu diesem Zweck wurde ein neues Paradigma eingeführt, das es ermöglichte, etablierte elektrophysiologische Korrelate der Inhibition in Verbindung mit numerischen Stimuli zu untersuchen. Die Ergebnisse belegen, dass die Inhibition der Wechselwirkung zwischen kontinuierlichen sensorischen und numerischen Informationen zugrunde liegt. Darüber hinaus zeigen Studie 2, dass das eingeführte Paradigma geeignet sein könnte, diese Prozesse über verschiedene Entwicklungsstadien und Rechenstufen hinweg zu untersuchen.

List of Original Articles

This dissertation is based on the following articles:

Pekár, J., & Kinder, A. (2020). The interplay between non-symbolic number and its continuous visual properties revisited: Effects of mixing trials of different types. Quarterly Journal of Experimental Psychology, 73(5), 698-710. https://doi.org/10.1177/1747021819891068

Pekár, J., Hofmann, W., Knakker, B., Tamm, S., & Kinder, A. (2023). Electrophysiological Correlates of the Interaction of Physical and Numerical Size in Symbolic Number Processing: Insights from a Novel Go/Nogo Numerical Stroop Task. Brain Sciences, 13(5), 702. https://doi.org/10.3390/brainsci13050702

Introduction

“Not everything that counts can be counted.
Not everything that can be counted counts."

Albert Einstein

Research in the field of numerical cognition has been occupying scientists for decades and from various disciplines such as psychology, neuroscience, anthropology, artificial intelligence and linguistics. Numbers indeed deserve great attention for many reasons. In order to become capable users of a wide range of cultural inventions that are related to numbers – e.g. simple arithmetic, geometry but also more complex ones such as architecture and engineering – children have to acquire language-based numerical skills and a symbolic number system. Anthropological findings illustrate best the importance and the relevance of this developmental milestone of numerical acquisition: cultures that lack numerical language also lack the ability to construct complex buildings. For example, the Mundurukú, an Amazonian indigene culture, lacks formal education and their language only entails number words from one to five. Analogically, they are not capable of building skyscrapers, airplanes or computers (Dehaene et al., 2006, 2008; Pica et al., 2004). The acquisition of numerical skills is important not only because of cultural and societal aspects but also because it closely affects the life of individuals. Approximately 3-6% of the population suffers from dyscalculia, an impairment in the development of number processing and calculation (Shalev, 2007; Shalev et al., 2000). Even though some may think that this problem can easily be aided by a simple calculator, studies show that individuals who suffer from this disorder are at higher risk for unemployment and various mental problems, such as depression. Providing help for these people is not only the prerequisite for giving them a chance to develop their full potential – our obligation that stems from the core value of liberal humanism – but it is also tied to major economic effects. Scientific results expand our understanding about the underlying components of mathematical and numerical cognition which in turn influences the law that allows these people to get the help and support they need. As such, studying numerical cognition is not autotelic but through legislative processes it also directly affects the life of individuals. The aim of the present dissertation is to widen our understanding about how numbers are processed in the brain and to open up new directions for further research.

Theoretical background

The study of numerical cognition is not limited to symbolic numbers and mathematics but it also includes continuous and discrete magnitudes. The main distinction between these two types of magnitudes is that continuous magnitudes, such as the length of a line, is not directly countable or measurable. Even if we are given unlimited time to figure out how long a given line is, we still need a measuring tape to measure the line length and the result will depend on the unit of measurement we chose to use, e.g. inches or centimetres. As opposed, discrete magnitudes, such as the number of dots in a dot array – in other words its numerosity – can be directly counted. When given enough time, we can directly enumerate the number of elements in a display or the number of tones in a sequence – provided that we are proficient users of a language that actually enables us to count that far, unlike the above mentioned Mundurukú who can only count up to five objects. Interestingly, discrete and continuous magnitudes are linked together and are inherently inseparable. For example, nine cookies are not only more cookies than five cookies but they also occupy a larger area on the plate. As such, continuous magnitudes are ubiquitous in nature and are also thought to serve as a basis for the acquisition of discrete quantities and symbolic numbers. This is referred to as the ‘symbol-grounding problem’, a central question in numerical research that focuses on how number words and number symbols acquire their meaning (Dehaene, 2001, 2007; Feigenson et al., 2013; Piazza, 2010; Stoianov, 2014 but see Leibovich & Ansari, 2016; Reynvoet & Sasanguie, 2016)

Numerical comparison tasks are one of the most widely used methods to investigate number processing. The task was first introduced by Moyer & Landauer (1967) who presented participants with two one-digit numbers. Their task was to decide as fast and as accurately as possible which one of the two Arabic numerals was numerically larger. Since then several variations of the paradigm have been implemented: (1) comparisons can be made between two numbers such as in the original task by Moyer & Landauer (1967) but numbers can also be compared to a standard, e.g. ‘Is this number larger than 5?’ (2) number pairs can be shown simultaneously or sequentially, and finally (3) quantities can be presented as symbolic numbers or number words such as 4 and four, canonical configurations such as the dot patterns on dice or they can be presented as non-canonical configurations like random dot patterns (Dehaene, 1996; Heine et al., 2013; Pinel et al., 2001; Szűcs & Csépe, 2005). Even though all possible variations of this task have probably been used in the past more than five decades since Moyer & Landauer (1967), they all yield one common finding: the well-established numerical distance or numerical ratio effect, i.e. faster reaction times and less errors are observed when comparing numerosities that are further from each other than when comparing numerosities that are closer to each other (e.g. 5 vs. 9 and 5 vs. 6, respectively). It has also been shown that participants decide faster and make less errors when comparing e.g. 5 and 10 than when comparing 50 and 55 despite the same distance between both number pairs. Therefore, the ability to discriminate between numbers and numerosities is better described with their ratio than with their absolute difference. These effects have been reported with non-symbolic as well as with symbolic numbers.

Numerous theories and models have been formulated in the field of numerical cognition. The most current and widely accepted one is the Triple Code Model (TCM) that was first described by Stanislas Dehaene (Dehaene, 1992; Klein et al., 2016; Moeller et al., 2015; Nieder, 2016; Nieder & Dehaene, 2009; Peters et al., 2016; Skagenholt et al., 2018). It is a neurocognitive model which postulates that numbers can mainly be stored in three different representational codes and each code has its own functionally distinct neural substrates and behavioural characteristics. First, symbolic numbers are stored in the visual Arabic number form that is primarily activated during calculations and parity judgement tasks. Second, the verbal phonological number form is activated during verbal counting and the retrieval of arithmetic facts, e.g. the elements of the multiplication table. Third, and probably most important, the analogue magnitude representation stores the semantic meaning of symbolic numbers and numbers words, i.e. the quantity a certain number represents, or in other words its cardinality. It is activated during subitizing, estimation, approximate calculations and numerical comparisons because during all of these tasks the quantity represented by a given number symbol or numerosity has to be accessed. Consequently, comparisons with symbolic as well as with non-symbolic numerosities activate the analogue magnitude representation (Dehaene, 1992; Dehaene et al., 1993; Dehaene & Changeux, 1993).

Research on humans and various animals has established the existence of an evolutionary ancient innate “number sense”, also referred to as the Approximate Number System (ANS). The ANS is responsible for the rapid and effortless extraction of large approximate numerosities from visual and other sensory scenes (Cantlon et al., 2009; Dehaene & Changeux, 1993; Feigenson et al., 2004; Gallistel & Gelman, 2000; Stoianov & Zorzi, 2012; Verguts & Fias, 2004). Pigeons, cats, fish – they all share the ability to select the larger one of two quantities, shall it be joining the larger group in order to ensure safety or selecting the larger portion of food in order to ensure adequate nutrition (Agrillo et al., 2008; Nieder et al., 2002; Watanabe, 1998). Single-cell studies in non-human primates found that numerosities are encoded by number selective neurons in a dedicated fronto-parietal network (for a review see Nieder, 2016). These neurons respond to preferred numerosities with specific tuning functions that emanate as partially overlapping Gaussian shaped activation patterns. For example, a dot display with four dots elicits the highest firing rate in the neuron which is tuned to four, but this neuron also responds to other numerosities – e.g. three and five – yet to a smaller degree. This activation pattern serves as the neurobiological foundation of a logarithmically compressed number scale. Neuroimaging studies identified the human counterpart of this network (Knops, 2017; Lyons et al., 2015; Piazza et al., 2004, 2007) and by now it is the widely accepted view that numbers and numerosities are represented along a spatially organized analogue number line with small numbers on the left and large numbers on the right side (Dehaene et al., 1993; Izard & Dehaene, 2008). Another similarity between numerical processing and the processing of other perceptual dimensions is that the Approximate Number System obeys Weber’s law. It is more difficult to distinguish numerosities that are closer to each other than numerosities that are further apart and that the minimal numerical difference that is still distinguishable increases as a function of numerosity (Dehaene, 2003; Piazza et al., 2004). This is the previously mentioned numerical distance or numerical ratio effect (Moyer & Landauer, 1967). The precision with which one can distinguish certain numerosities is called the ANS acuity which can be formally quantified as the Weber fraction (DeWind et al., 2015). It is assumed that the precision of the ANS predicts symbolic mathematical abilities and that the ANS underlies the development of arithmetical and mathematical capacities. During typical development, as children acquire numerical language skills and a language-based symbolic number system, numerical symbols (number words and symbolic numbers) are mapped onto the ANS (Dehaene, 2001, 2007; Feigenson et al., 2013; Piazza, 2010; Stoianov, 2014 but see Leibovich & Ansari, 2016; Reynvoet & Sasanguie, 2016). Thus, number-selective circuits in the bilateral intraparietal sulci (IPS) that host the ANS also act as primary semantic representation of symbolic number that are activated during number comparison tasks.

A major question in numerical research when studying the ANS is how continuous magnitudes and non-symbolic numerosities interact (for reviews see Gebuis et al., 2016; Leibovich, Katzin, et al., 2016). It has been pointed out that discreet and continuous magnitudes are linked together and are inherently inseparable. This feature of continuous and discrete magnitudes is apparent in several real-life situations as well. For example, when we try to estimate which queue will move faster at the airport security or at the grocery store. In these situations, we usually do not start counting the people in the line, but rather we rely on visual cues, such as the length of the queue and how closely people are standing next to each other. Thus, we arrive at a fair estimate and comparison based on the perceptual cues at hand.

What role do continuous magnitudes play in numerical processing or, to put it differently, how do the sensory characteristics of numerical stimuli influence numerical comparisons and judgements? The interaction of sensory cues and numerical information is usually investigated with tasks that are modified versions of the original numerical comparison task that was introduced by Moyer & Landauer (1967). A common characteristic of these experiments is that two stimulus dimensions are concurrently manipulated – i.e. the continuous features and the numerical information – in order to create congruent and incongruent trials. By comparing these two kinds of trials it is possible to draw inferences about the interaction between continuous and discreet magnitudes.

In non-symbolic numerosity processing the dot comparison task is most common method to investigate the influence of continuous magnitudes. In this version of the numerical comparison task stimuli are visual dot displays with varying number of dots. The numerical information, i.e. the number of the displayed dots, however, is not the sole dimension of the stimuli that is manipulated. Dot diameter, dot density, aggregate dot surface and convex hull (i.e. the smallest contour around the dots) – they can all be either congruent or incongruent with numerical information (please see Figure 1). Thus, to create congruent and incongruent trials, these sensory properties are manipulated in a way that they may or may not comply with natural contingencies. If we replace dots with cookies, we get a very illustrative everyday life example. As mentioned before, nine cookies are not only more cookies than five cookies, but they also occupy a larger area on the plate, i.e. they have a larger convex hull, a larger aggregate surface and they have to be placed more densely on the plate. In this sense, '*larger is more*' constitutes the crucial natural contingency to be considered when creating congruent and incongruent trials. It follows that on congruent trials the more numerous dot array also has larger visual properties while on an incongruent trial the array with more dots has smaller visual cues. It is also important to note, that researchers have put considerable efforts into developing several different methods to control the relationship between sensory cues and numerosity when creating such dot patterns (De Marco & Cutini, 2020; Dehaene et al., 2005; Gebuis & Reynvoet, 2011; Guillaume et al., 2020; Salti et al., 2016). Depending on the research question at hand, participants are asked to select the more numerous dot array or the array with larger visual characteristics. Several studies have been conducted to investigate the relationship between sensory cues and numerosity. For example, Gebuis & Reynvoet (2012a) showed that when the continuous magnitudes were manipulated in a certain way, performance almost reached ceiling effect. However, when the same continuous magnitudes were manipulated in a different way – while the numerosity information was kept constant – performance dropped close to chance level. In sum, studies that investigated the interplay between continuous visual properties and non-symbolic numerosity have confirmed that sensory cues have an immense effect on numerical judgements.

The notion that numerosity and continuous magnitudes interact seems to be supported not only by behavioural investigations but also by single cell and human neuroimaging studies (Dormal & Pesenti, 2009; Kadosh et al., 2005; Nieder, 2016; Pinel et al., 2004; Tudusciuc & Nieder, 2009). It has been shown that certain neurons in the primate brain are either tuned to discrete numerosity or to continuous magnitudes or they encode both discrete and continuous stimulus features. These three types of neurons were found to be intermingled in the monkey parietal cortex. Analogically, human neuroimaging studies have also showed segregated but also overlapping activation patterns in response to

Figure 1: Nine cookies on a plate are not only more cookies than five cookies, but they also have a larger convex hull (larger contour around them), a larger aggregate surface and a larger density as they have to be placed closer to each other on the plate.[1]

continuous magnitude and numerosity in the human parietal cortex. Thus, it is possible that the interaction between continuous magnitudes and numerosity arises on the basic neuronal level. In line with this hypothesis, Eiselt & Nieder (2013) reported that certain population of neurons responded when monkeys had to select the 'more numerous' and also when they had to select the 'larger' stimulus. Another population of neurons responded when the monkeys had to select the 'less numerous' and also when they had to select the 'smaller' stimulus. In other words, the congruency effect has been shown on the basic neuronal level meaning that there are neurons that are involved in both 'greater' and 'larger' responses while other neurons are involved in 'fewer' and 'smaller' responses. Thus, these findings together might explain the interplay between continuous magnitudes and numerical information.

[1] Graphic designed by Marina González Gómez

There are two more processes that may mediate the relationship between continuous and discrete magnitudes: integration and inhibition (Gebuis et al., 2016; Gevers et al., 2016; Leibovich, Kallai, et al., 2016; Leibovich, Katzin, et al., 2016). Both ideas have been linked to the famous conservation error that was described by Jean Piaget in 1954 (Piaget, 1954). He observed that children between the age of 18 months and 6/7 years – the age he named the preoperational stage – tend to err when they are asked which one of the two rows of coins is more numerous. Interestingly, they can only tell the correct answer if the line length correlates positively with the number of coins. However, when the experimenter moves the coins to create a negative correlation between line length and coin number, the children say the longer row is more numerous. Thus, they fail to conserve number. The notion of integration and inhibition provide different explanations for this phenomenon.

The inhibition account proposes that both continuous and numerical information are processed when making numerical judgements but there is an inherent bias towards continuous magnitudes (Leibovich, Kallai, et al., 2016; Leibovich, Katzin, et al., 2016). Children cannot solve the conservation task correctly because the required inhibitory skills haven’t fully developed yet. Consequently, they completely rely on the continuous magnitudes. This idea is supported by studies that investigated the role of cognitive control processes in numerical cognition during different developmental stages. For example, in a study by Houdé et al. (2011), so called conservers (9-10 year-old school-aged children) and non-conservers (5-6 year-old pre-schoolers) were asked to solve a Piaget-like conservation task. Behavioural results showed that the 9-10 year-old but not the 5-6 year-old children could correctly solve the task. Neuroimaging results revealed differences in brain activity in a fronto-parietal network, in particular in the inferior frontal gyrus, between those who failed and those who succeeded in the task. The authors interpreted these results as the activation of inhibitory processes that are necessary to correctly solve the task. In sum, the inhibition account views continuous magnitudes as a source of interference and as such, they should be inhibited. Any effect of continuous magnitude stems from insufficient inhibition of visual properties not only in children but also in adults. This idea is further supported by behavioural investigations that showed that performance in dot comparison tasks increases with age. This increase is, however, modulated by better performance on incongruent trials, i.e. on trials where the visual characteristics have to be inhibited to correctly solve the comparison (Szűcs, Nobes, et al., 2013).

The idea of integration stems from the Sensory Integration Theory which was first formulated by Gebuis & Reynvoet (2012a), who implemented a novel method to directly investigate the role of continuous magnitudes in a dot comparison task (Gebuis & Reynvoet, 2011, 2012a; Gevers et al., 2016). They manipulated the continuous magnitudes of dot arrays to create different congruency conditions and compared performance between trials that had the same numerosity information but differed in terms of congruency and visual cue manipulation. Their pattern of results was consistent with the notion of sensory integration according to which continuous magnitudes are weighed and integrated together in order to arrive at a numerosity estimate when comparing large approximate numerosities. As mentioned before, when creating dot patterns, there is always a visual cue that correlates either negatively or positively with numerosity. The relationship between the different sensory cues and their relationship to numerosity is essential during the so-called integration procedure. Every visual cue gets a weight. The weight assigned to a specific visual cue depends on its saliency. That is, if two dot arrays differ to a larger extent in terms of convex hull than in terms of aggregate surface, then convex hull gets a larger weight. According to the Sensory Integration Theory, children fail to solve to conservation task because they have not yet learnt the relationship between numerosity, line length and coin density. It is important to note, that this account is essentially a perceptual one. However, it goes beyond to what findings of single cell studies would suggest because it proposes a complex influence of continuous magnitudes on numerical judgements. Not only the relationship between numerosity and a specific sensory cue but also the relationship between the various sensory cues influences numerical judgements.

These theories have been formulated in the realm of non-symbolic numbers. However, symbolic number processing is also affected by the sensory characteristics of the stimuli. This phenomenon is investigated with the so called the numerical Stroop task. In this version of the numerical comparison task stimuli are pairs of Arabic numbers that are manipulated in terms of physical size and numerical size in order to create congruent, neutral and incongruent trials (Henik & Tzelgov, 1982). On congruent trials the number that is numerically larger is also larger in physical size (4 5). Analogically, on incongruent trials, the numerically larger number is smaller in physical size (4 5) while on neutral trials both digits have the same physical and/or numerical size (4 5 or 4 4). Again, depending on the research question at hand, participants may be asked to decide with button press which number is numerically larger or physically larger. Studies that implemented this paradigm have repeatedly confirmed that comparisons with symbolic numbers are also affected by the sensory cues, i.e. the physical size of the numbers (Gebuis et al., 2010; Henik & Tzelgov, 1982; Kaufmann et al., 2005, 2008; Szűcs & Soltész, 2007, 2012; Tang et al., 2006). Besides the well-established numerical distance effect, i.e. less errors and shorter reaction times on large compared to small numerical distances, also facilitation and interference effects can be observed. When comparing congruent trials participants make less errors and decide faster which digit is larger than when comparing neutral and incongruent trials. Also, more errors and longer reaction times are observed on incongruent than on neutral and congruent trials. Interestingly – unlike in the classical Stroop paradigm – in the numerical Stroop task the facilitatory and interference effects are bidirectional. Not only numerical size judgements are affected by the physical size of the numbers but also the other way around. Physical size judgements are also affected by the numerical size of the digits. Thus, the paradigm has been an excellent tool for investigating whether symbolic numbers and other continuous magnitudes are processed by separate systems or by a common magnitude system (Gebuis et al., 2010; Henik & Tzelgov, 1982; Huang et al., 2021; Kadosh et al., 2007; Soltész et al., 2011; Szűcs & Soltész, 2007, 2012; Tang et al., 2006; Yao et al., 2015). Two opposing accounts, the *early interaction account* and the *late interaction account*, have been postulated to explain the interaction of numerical and non-numerical parameters in the numerical Stroop task. Researchers implement neuroimaging and electrophysiological studies in order to differentiate between these two accounts by looking at the locus of the interaction between physical size and numerical size. According to the *early interaction* account the numerical and physical dimensions of the stimuli are processed by a common magnitude system and therefore the interaction can be observed at an early stage, prior to motor response selection. As opposed, the *late interaction* account suggests that the physical and the numerical dimensions are processed parallel by distinct neural substrates and the interaction occurs at a later stage, upon response initiation. In electrophysiological investigations event-related potential components are used to search for the locus of interaction while implementing the numerical Stroop task. These electrophysiological markers include the P3 and the lateralized readiness potential (LRP). The P3 a positive going deflection usually observed between 300-500 ms after stimulus presentation over centro-parietal electrodes. It is thought to reflect stimulus evaluation and categorization processes. The LRP is an electrophysiological marker of selective motor activation and it is calculated by subtracting ipsilateral activity from contralateral activity of the response hand. While congruency effects on the P3 component suggest that the interaction of physical and numerical size information occurs at the stimulus evaluation and categorization stage, congruency effects on the LRP suggest that the interaction occurs at the motor response selection stage. Taken together, interaction effects on the P3 support the early interaction account and a shared neural substrate while interaction effects on the LRP support the late interaction account and distinct neural substrates (Gebuis et al., 2010; Kadosh et al., 2007).

The matter of shared versus distinct neural representation of magnitude has been a subject of vigorous debate. For example, investigating the P3 and the LRP, Gebuis et al. (2010) reported effects that favour the early interaction account while Kadosh et al. (2007) also found evidence for a late interaction of physical and numerical size. Thus, there is no consensus in this matter. However, it is also important to note, that no study so far has gone beyond searching for the locus of the interaction and investigated the nature of the psychological processes underlying the numerical Stroop effect. Analogically to non-symbolic numbers, where inhibition has been named as important factor in the interplay between continuous sensory cues and numerosity, the interaction of physical and numerical size may be explained by inhibitory processes. This idea is supported by behavioural studies showing that executive functions are implicated in the development of mathematical abilities. Only a handful of papers that implemented the numerical Stroop task investigated inhibitory processes but their main goal was to unravel certain developmental profiles or to elucidate the functional definition of certain electrophysiological markers (Bryce et al., 2011; Szűcs & Soltész, 2012). Thus, those studies did not investigate whether inhibition underlies the interaction of physical size and numerical size in the numerical Stroop task.

The present work consists of two studies both of which aim to investigate the exact nature of the interaction between continuous sensory properties and numerical information. While the first study implemented a dot comparison task to examine the combined effect of different continuous magnitudes on non-symbolic numerical comparisons, the second study turned to symbolic number processing and introduced a modified version of the numerical Stroop task to investigate the psychological processes, inhibitory processes in particular, that may underlie the interaction of physical size and numerical size of Arabic number pairs.

Study 1

The first study relates to the original work by Gebuis & Reynvoet (2012a) who introduced a novel method to manipulate the relationship between continuous magnitudes and numerical information in a dot comparison task. Their method is a special one because it enables to concurrently manipulate various visual cues either congruently or incongruently with numerosity within one trial. For example, the more numerous dot array may have a larger convex hull and larger sized dots (congruent & congruent) or it may have a larger convex hull but smaller sized dots (congruent & incongruent, for more information please see Methods section of Study 1). Using this method, Gebuis & Reynvoet (2012a) found a positive relationship between convex hull and numerosity, i.e. better performance when the more numerous dot display had a larger convex hull. However, they found a negative relationship between dot diameter and numerosity. Performance was better when the more numerous dot array had smaller dots. In other words, performance was better when convex hull adhered to the '*larger is more*' natural contingency but in the case of dot diameter performance was better when this natural contingency was violated. Finding this pattern of congruency effect is a curious result because if the interaction between continuous magnitudes occurs on a basic neuronal level, then the effect should be consistent across continuous magnitudes. As mentioned before, monkey single cell studies have demonstrated the congruency effect on the basic neuronal level, meaning that there are cells that are involved in both *larger* and *more* responses while other cells are involved in *smaller* and *fewer* responses but not in *smaller* and *more* responses (Eiselt & Nieder, 2013). Analogically, performance should be better when the *more numerous* dot array has *larger* and *not smaller* dot diameter. Thus, Study 1 aimed to assess the stability and the robustness of the reversed diameter effect as well as the pattern of congruency effects as a whole across the different visual cue manipulation methods that Gebuis & Reynvoet (2012a) implemented in the original study. To this end, trials of different types were presented in a mixed fashion and compared to the original version of the task, where different trial types were presented in separate blocks. The study replicated the reversed diameter effect and it also showed that the influence of continuous magnitudes can be altered by changes in stimulus history, i.e. the influence of sensory cues increased when different trial types were shown in a mixed fashion. Consequently, even if there is an interaction between continuous magnitudes and numerosity on the basic neuronal level, it cannot fully explain the pattern of the observed behavioural results because it does not account for either the reversed diameter effect or for the increased reliance on visual cues when the stimulus history is changed. In sum, these effects together imply a more complex influence of continuous magnitudes on numerical judgements and that the interaction between continuous sensory cues and numerical information can be due to either inhibition or integration.

Study 2

Study 2 turned to symbolic numbers and investigated whether inhibition underlies the interaction of physical and numerical size which occurs in the numerical Stroop task. As mentioned before, the numerical Strop paradigm has been used to look for the locus of the interaction between physical and numerical size, and therefore to assess whether these two stimulus dimensions are processed by shared or distinct neural substrates (early versus late interaction). However, no study has so far directly assessed what psychological processes underlie this interaction. In Study 2, in order to investigate whether inhibition underlies the observed interaction effects in the numerical Stroop task, I introduced a novel hybrid paradigm and combined the numerical Stroop task with a go/nogo task. This modification allowed to investigate the well-established electrophysiological correlates of inhibition and assess whether and how these components are affected by physical size manipulation in the numerical Stroop task.

In inhibition research the most commonly measured electrophysiological correlates of inhibition are the N3 and the P3 components. They are usually observed in go/nogo tasks (Bruin et al., 2001; Bruin & Wijers, 2002; Falkenstein et al., 1999) where participants are presented with a stream of simple stimuli and are asked to respond with button press to frequent go trials (e.g. the letter M) and withhold from responding on infrequent nogo trials (e.g. the letter 'W'). In these tasks the N2 component is a fronto-central stimulus locked negative going event-related potential that occurs between 200-350 ms post-stimulus and it is more negative on nogo than on go trials. The P3 ERP component is a positive going stimulus locked deflection appearing between 300-500 msec which is more positive on nogo than on go trials and – as opposed to the previously mentioned categorization-related P3 components – the P3 in inhibition tasks has a more anterior topography. Importantly, these components have also been shown to be sensitive to interference from irrelevant stimulus dimensions (Brydges et al., 2012, 2013; Groom & Cragg, 2015; Xie et al., 2017).

The novel go/nogo numerical Stroop task that I implemented in Study 2 was also designed to measure the inhibition-related N2 and P3 components and to assess whether and how they are affected by the interference from the irrelevant stimulus dimension of physical size in the numerical Stroop task. Participants were presented with Arabic number pairs of various numerical distances where the physical and numerical size of the numbers were manipulated in order to create congruent, neutral and incongruent trials. Importantly, the response mode was additionally manipulated in order to create go and nogo trials: participants were instructed to press a button when the number on the one side was numerically larger and withhold from pressing the button when the number on the other side was numerically larger.

Analysing congruency and go/nogo effects while measuring the inhibition-related N2 and P3 ERP components allowed to directly address the role of inhibition in the interaction of physical and numerical size and also to gain insights about the processing stage at which inhibition may occur. Similar to the previously mentioned early and late interaction accounts, if inhibition is implicated in the numerical Stroop effect, then it may occur at an early stage, i.e. the interference from irrelevant physical size is inhibited very early, presumable at the stage of perceptual processing. Or inhibition may occur at a later stage, possibly at the stage of response initiation. Since the paradigm included not only congruent and incongruent but also neutral trials, it was possible to assess whether the facilitatory effect of physical size are also observed after introducing the go/nogo manipulation to the numerical Stroop task. In sum, the findings of this study confirmed that interference effects in the numerical Stroop are probably resolved by inhibitory processes at a very early stage between 200-350 ms. Furthermore, facilitatory effects were only found on go trials which means facilitation is affected by different cognitive control processes than those required by go versus nogo trials.

In the next sections these two studies are described in detail as they were published and submitted for peer-review (Study 1 & Study 2, respectively). The last part of this work summarizes the main findings of both studies and discusses important implications.

Study 1:

The interplay between non-symbolic number and its continuous visual properties revisited: Effects of mixing trials of different types

Judit Pekár & Annette Kinder

Pekár, J., & Kinder, A. (2020). The interplay between non-symbolic number and its continuous visual properties revisited: Effects of mixing trials of different types. Quarterly Journal of Experimental Psychology, 73(5), 698-710. https://doi.org/10.1177/1747021819891068

1. Introduction

The concept of numerosity is essential in everyday life. For example, we estimate and compare numerosities when we try to choose the shorter queue in the supermarket or when we try to pick the plate with more biscuits in the cafeteria. The most prominent theory in numerical cognition is the Approximate Number System (ANS). According to this theory, humans and animals share an evolutionary ancient system which is responsible for the rapid and effortless extraction of numerosities from visual and other sensory scenes (Cantlon et al., 2009; Feigenson et al., 2004). In particular, single-cell studies have found that number selective neurons in a dedicated fronto-parietal cortical network encode numerosity in the primate brain (for a review see Nieder, 2016). Furthermore, neuroimaging studies support the existence of the human counterpart of this network (Knops, 2017; Lyons et al., 2015; Piazza et al., 2004, 2007). While behavioural, neuroimaging, neurophysiological and computational studies support the existence of the ANS (Cantlon et al., 2009; Dehaene & Changeux, 1993; Feigenson et al., 2004; Gallistel & Gelman, 2000; Stoianov & Zorzi, 2012; Verguts & Fias, 2004), its exact mechanisms are less clear. In particular, it is still an open question whether and how sensory cues influence numerosity processing (Leibovich, Katzin, et al., 2016).

Classic theories of the ANS suggest that numerosity processing occurs independently of continuous sensory properties present in the stimuli, i.e. our estimate about the biscuits in the cafeteria is not biased by their size or by how densely they are placed on the plate. Discrete numerosity, however, is always confounded by continuous magnitudes: considering the example above, the total volume of the biscuits will be larger on the plate with more pieces, provided that all biscuits are of the same size. However, when the total biscuit volume on both plates are equal, then the size of the biscuits must be smaller on the plate with more biscuits. Even though researchers have put considerable effort into developing methods that can control the relationship between numerosity and its confounding sensory properties (Dehaene et al., 2005; Gebuis & Reynvoet, 2011; Salti et al., 2016), there is always a continuous magnitude which correlates – either positively or negatively – with numerosity (see for example Smets, Moors, & Reynvoet, 2016). Thus, inspired by the ubiquitous nature of sensory properties, in the last few years, attention has been shifted to investigating what role non-numerical parameters play in numerosity processing. Based on these studies, more recent theories postulate that continuous properties influence or might even explain the processes underlying the estimation and comparison of large approximate numerosities (Clayton et al., 2019; Gebuis et al., 2016; Gevers et al., 2016; Leibovich, Kallai, et al., 2016; Leibovich, Katzin, et al., 2016).

Thus, on the one hand, numerous studies support pure number sense (Cantlon et al., 2009; Feigenson et al., 2004; Gallistel & Gelman, 2000; Piazza et al., 2007; Stoianov & Zorzi, 2012; Verguts & Fias, 2004) while a growing number of behavioural, neurophysiological and neuroimaging studies suggests that continuous sensory properties play an important role in numerosity processing (Gebuis & Reynvoet, 2012a, 2012b, 2013; Leibovich & Ansari, 2017; Leibovich & Henik, 2014; Salti et al., 2016; Smets et al., 2015; Soltész & Szűcs, 2014). Usually, in these experiments participants have to solve a dot comparison task in which two dot arrays of varying numerosities are presented to them. They are asked to indicate whether the first or the second array had contained more dots. Unbeknownst to the participants, continuous magnitudes are manipulated to create different congruency conditions. For example, on a congruent trial the array with more dots has larger visual cues, i.e. larger sized dots or a larger convex hull (larger contour around the dot array). In contrast, on an incongruent trial, the more numerous dot array has smaller visual characteristics, i.e. smaller dots or a smaller convex hull. With this method, one can manipulate several continuous magnitudes together in the same direction, e.g. dot size and convex hull are both larger on the array with more dots (congruent & congruent), or they are both smaller (incongruent & incongruent). The continuous magnitudes can also be manipulated in the opposite directions, e.g. larger dot size and a smaller convex hull on the array with more dots (congruent & incongruent). It is even possible to keep one continuous magnitude constant, while making the other congruent or incongruent. For example, the two dot arrays have the same convex hull but the more numerous one has larger dots (equated & congruent) or the less numerous one has larger dots (equated & incongruent, see also Figure 1). Comparing performance between trials with the same numerosity information but different congruency and visual cue manipulation, gives us insight into the effects of continuous magnitudes on numerosity processing. Findings of experiments which implemented this kind of sensory cue manipulation have raised the attention to the relationship between numerosity and its continuous magnitudes and questioned the existence of a pure number sense which is independent of continuous magnitudes (Gebuis & Reynvoet, 2012a, 2012b; Salti et al., 2016; Smets et al., 2015).

Sensory Integration Theory provides an alternative approach to the Approximate Number System (Gebuis et al., 2016; Gevers et al., 2016). Proponents of this theory claim that sensory cues are weighed and integrated during numerosity processing. Rather than deriving numerosity independently of sensory cues, during the so called integration procedure multiple sensory cues compete with each other in their weight given to the numerosity estimate. For example, if there is a stimulus pair which differs to a larger

Figure 1

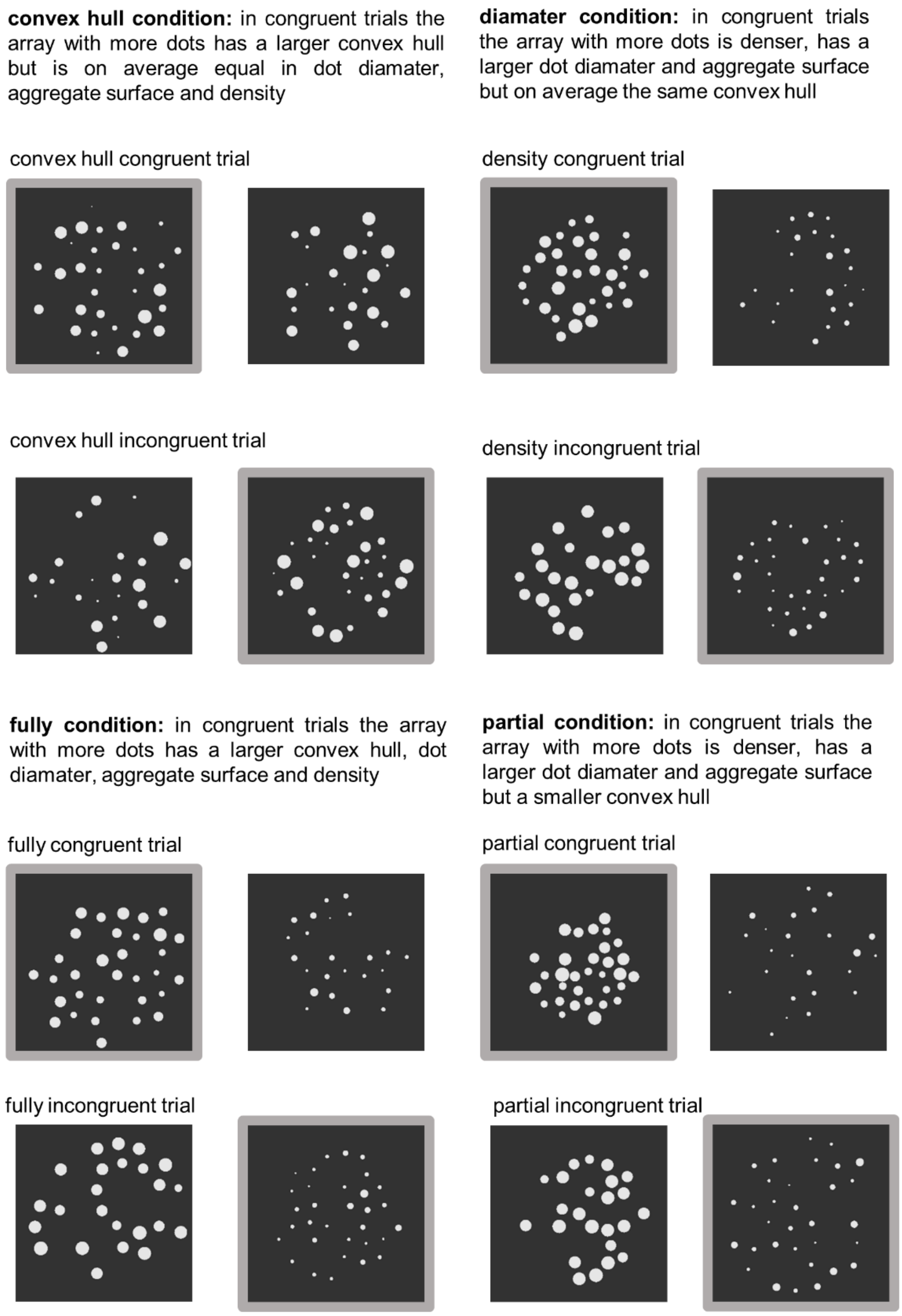


Figure 1: Examples of congruent and incongruent trials. For each of the four conditions one congruent and one incongruent stimulus pair is shown. The more numerous stimulus is marked with grey border.

extent in terms of convex hull than in terms of dot diameter size, more weight might be given to the convex hull. This mechanism can explain the pattern of congruency effects found by Gebuis & Reynvoet (2012a). Specifically, they implemented the visual cue manipulation method described above and created four different visual cue conditions. In two conditions, only a single visual cue (convex hull) or a set of visual cues (average dot diameter, aggregate surface of the dots and density, from now on referred to as diameter) were manipulated and they were either congruent or incongruent with numerosity. In the two other conditions, both visual cues were manipulated together, although in different directions (see Methods section for more detail). Consequently, neither convex hull nor any other visual cues were consistently informative since they did not correlate with numerosity throughout the task. Interestingly, Gebuis & Reynvoet (2012a) found opposite congruency effects in those two conditions, in which a single visual cue was manipulated. In the convex hull condition, performance was better when more dots occupied a larger area (other visual cues were equated). In the diameter condition the congruency effect reversed. Namely, performance was better when the array with more dots had a smaller average diameter, smaller aggregate surface and a smaller density (convex hull was equated). When all visual cues were manipulated in the same direction, the opposite effect of convex hull and the other sensory cues cancelled each other out, resulting in the absence of a congruency effect. However, when convex hull and diameter were manipulated in the opposite direction, their differential effect on the performance led to an augmented congruency effect. The authors argued that this pattern of findings support the idea that participants integrate various types of continuous dimensions, possibly by means of an additive weighing process, even when they are uninformative about numerosity.

Single cell studies suggest that neurons in the primate brain are either tuned to discrete numerosity or continuous magnitude, or they encode both discrete and continuous stimulus features. Moreover, these three types of neurons were found to be intermingled in the monkey parietal cortex (for example Tudusciuc & Nieder, 2009, for an overview see Nieder, 2016). These findings are in line with human neuroimaging studies which showed overlapping but also segregated activations in response to continuous magnitude and numerosity in the human parietal cortex (Dormal & Pesenti, 2009; Kadosh et al., 2005; Pinel et al., 2004). So the notion that numerosity and continuous magnitude are integrated seems to be supported by single cell and human neuroimaging studies. However, if there is an interaction between numerosity and continuous magnitudes on the basic neuronal level, then this interaction should be consistent across different continuous magnitudes. That is, some cells should respond to larger numerosity and larger convex hull while other cells should respond to larger numerosity as well as larger diameter and not smaller diameter. In line with this hypothesis, certain neurons in the monkey prefrontal cortex responded not only when monkeys had to select the more numerous but also when they had to select the larger stimulus. Another group of neurons responded when monkeys had to select the less numerous and the smaller stimulus. Thus, the congruency effect has been shown at a basic neural level, meaning that there are cells which are involved in both “greater” and “larger” responses and other cells that are involved in both “fewer” and “smaller” responses (Eiselt & Nieder, 2013). Taken together, these findings might provide an explanation of the interference effect between numerosity and continuous magnitudes found in behavioural investigations (Gebuis et al., 2010; Szűcs & Soltész, 2007). However, the reversed congruency effect in the diameter condition found by Gebuis & Reynvoet (2012a), is difficult to reconcile with these findings because they suggest that displays with larger dot diameter should be judged as more numerous.

Further studies, which address the role of continuous magnitudes on numerosity processing, may add to our understanding about the interplay between discrete numerosity and continuous sensory cues. For example, it has been shown that the effect of continuous magnitudes on dot comparison performance can be altered by changes in instructions, task difficulty, stimulus duration and task context (Leibovich et al., 2015; Leibovich-Raveh et al., 2018). Thus, the weights that are given to continuous magnitudes are not static but seem to depend on various experimental factors. As some of these factors (such as presentation time) affect bottom-up processing whereas other factors (such as instructions) affect top-down processing, weights seem to depend on both top-down and bottom-up processes as well as on their interaction. As a result, sensory cues are processed in a highly flexible and adaptive fashion. Despite this growing body of research, it is still an open question, how exactly the weights of the continuous magnitudes are adjusted. For example, it is unclear whether the weights are adjusted within trials depending mainly on the stimulus features in the current trial or the weights are rather adapted gradually across many trials. The latter idea is supported by a study by Odic et al. who found that performance on a dot comparison task was influenced by trial order (Odic et al., 2014). Participants who were presented with easy trials first performed better on the task than participants who received difficult trials first. Thus, comparison performance can be altered through the history of previous trials.

If the weighing process indeed depends on trial history, it should make a difference whether trials of the same type are presented consecutively (blocked presentation) or whether different trials types are mixed together (mixed presentation). In the original study, Gebuis & Reynvoet (2012a) manipulated the sensory characteristics of the two dot arrays in a block-wise fashion, hence trials of the same type were shown consecutively to the participants. For example, in a certain block, the dot diameter of stimulus pairs was always manipulated in the same way (either congruently or incongruently to numerosity) whereas convex hull remained constant. In another block, it was the opposite: the convex hull of stimulus pairs was manipulated in the same way (congruent or incongruent), while dot diameter remained constant. But, assuming that the weighing process is flexible and adaptive as suggested by Leibovich et al., and also assuming that trial history has an effect on dot comparison performance as suggested by Odic et al., it is possible that the blocked presentation mode influences the weights given to the continuous magnitudes in a particular way (Leibovich et al., 2015; Leibovich-Raveh et al., 2018; Odic et al., 2014). For example, if during the entire block the two stimuli in a trial differ with regard to convex hull whereas the other continuous cues are kept constant, the weight assigned to convex hull may increase or decrease compared to the weights of other continuous magnitudes. Hence, the specific pattern of congruency effect may at least partially be accounted for by the block-wise presentation mode. It seems even possible that the reversed congruency effect (that corresponds to a negative weight) in the diameter condition is merely a result of the blocked presentation mode and that it is not found under altered circumstances, e.g. when different trial types are mixed together. Therefore, we would like to investigate whether changing the trial history can alter the interaction between continuous magnitudes and numerosity and thus change the pattern of congruency effects reported by Gebuis & Reynvoet (2012a).

In sum, our study has two aims. First, we examine whether changing the trial history by mixing trials of different types can alter the pattern of congruency effects shown by Gebuis & Reynvoet (2012a). Second, we are particularly interested in the reversed congruency effect in the diameter condition because this effect is difficult to reconcile with single-cell studies which would suggest the opposite. Thus, we would like to investigate whether this effect remains stable when the presentation method is changed from blocked to mixed. To this end, we presented participants with different trials types in a mixed fashion and compared it with the exact replication of the original study (Gebuis & Reynvoet, 2012a).

2. Methods

2.1 Participants

Data were collected from 34 individuals (6 males, age: M=26.69, SD=5.96, range: 18 years 9 months – 36 years 5 months). All participants had normal or corrected-to-normal vision. They gave written informed consent and received course credit for their participation. Three participants (all of them females) were excluded from the data analysis because of low performance on the Dyscalculia Screener (Butterworth, 2003, see below). In total, data from 31 participants were analysed (6 males, age: M=26.26, SD=5.99, range: 18 years 9 months – 36 years 5 months).

Gebuis and Reynvoet (2012a) reported a congruency effect for convex hull with an effect size of *Cohen's* d_z=1.54 and a reversed congruency effect for diameter with an effect size of *Cohen's* d_z=.86 (calculated as *Cohen's* $d_z = t/\sqrt{n}$, according to Rosenthal, 1991). Power to find these effects with N=31 subjects (α = .05) were $1\text{-}\beta$ = 1 and .99, respectively (Faul et al., 2007). A sample size of N=31 allows to detect possible moderations of congruency effects by presentation mode with an effect size of *Cohen's* d_z =.52 (i.e. medium sized-effect according to Cohen, 1988).

2.2 Stimuli and Tasks

2.2.1. Dot comparison Task. Stimuli were pairs of dot arrays which were presented consecutively to the participants. Their task was to indicate with a button press whether the first or the second image had contained more dots. Dot arrays were constructed the same way as those used by Gebuis & Reynvoet (2012a). White dots were presented on a dark grey background, dot size ranged from 0.11 degrees to 0.79 degrees in visual angle. Four visual properties were manipulated which are thought to influence numerosity judgements: (1) *convex hull* (area within the smallest contour around the dot array), (2) *aggregate surface of the dots*, (3) *average dot diameter* and (4) *density* (aggregate surface divided by convex hull).

It is important to note that three of these visual properties, namely aggregate surface, average diameter and density, are highly related. Therefore, it was not possible to differentiate between them. For example, if average dot diameter is increased, aggregate surface and density will also automatically increase while convex hull can remain constant. For this reason, aggregate surface, average dot diameter and density were manipulated together in one condition, which will be referred to as *diameter in/congruent condition* hereafter.

The manipulated visual cues could either be congruent or incongruent with numerosity. A stimulus was considered congruent if the visual property in question was greater for the array with more dots. In the *(1) convex hull in/congruent condition* the visual property manipulated was the area within the smallest contour around the dot array whereas density, aggregate surface and average diameter were kept equal across numerosities and congruency conditions. Thus, on *convex hull congruent trials* the array with more dots was associated with greater convex hull around the dot array. In contrast, on *convex hull incongruent trials*, arrays containing a larger number of dots had a smaller convex hull. Density, aggregate surface and average diameter were kept constant in this condition. That is, congruent and incongruent trials did not differ from each other with regard to these properties. In the *(2) diameter in/congruent condition* the convex hull was kept constant while the density, the average diameter of the dots and their aggregate surface were either congruent or incongruent with numerosity. In the *(3) fully in/congruent condition* both diameter and convex hull were manipulated together and they could be either congruent or incongruent with numerosity. As a result, on *fully congruent trials* all visual cues were larger in stimuli containing more dots. In contrast, on *fully incongruent trials* the more numerous dot arrays were associated with smaller visual cues. In the *(4) partially in/congruent condition* diameter and convex hull were manipulated in opposite ways: when convex hull was incongruent, diameter, density and aggregate surface were congruent and vice versa. So in short, *partially congruent trials* were congruent in diameter (as well as related visual cues) and incongruent in convex hull. Whereas, *partially incongruent trials* were incongruent in diameter (as well as related visual cues) and congruent in convex hull.

Each condition contained 192 trials half of which were congruent and the other half were incongruent (96 trials each). Every trial consisted of two dot arrays presented consecutively. One of the dot arrays always contained 24 dots, the other dot array could contain 16, 18, 20, 29, 32, 36 dots resulting in six different numerosity combinations equivalent to three different ratios. The ratios were 1:2, 1:3, 1:5 calculated as *(larger* *number – smaller number)/smaller number*. Trials within each condition were counterbalanced with respect to congruency, numerosity combinations and the number of dots in the first array.

It is important to note that as in the original study, the visual cues were not informative of numerosity, neither within a single cue manipulation condition nor throughout the whole task. Within each visual cue manipulation condition, half of the trials were congruent and the other half were incongruent. Moreover, the differences in continuous magnitudes and numerosity between stimulus pairs did not significantly correlate ($R<0.1$, $p>0.05$). Only in the *convex hull in/congruent condition* the aggregate surface was informative, since it was always larger for the more numerous dot array. It was not possible to control for this visual cue without making other cues informative. However, the difference in aggregate surface occurred in the same direction on congruent and incongruent trials. So any difference between trials with different congruency cannot be explained by participants utilizing aggregate surface as a cue to infer numerosity (Figure 2).

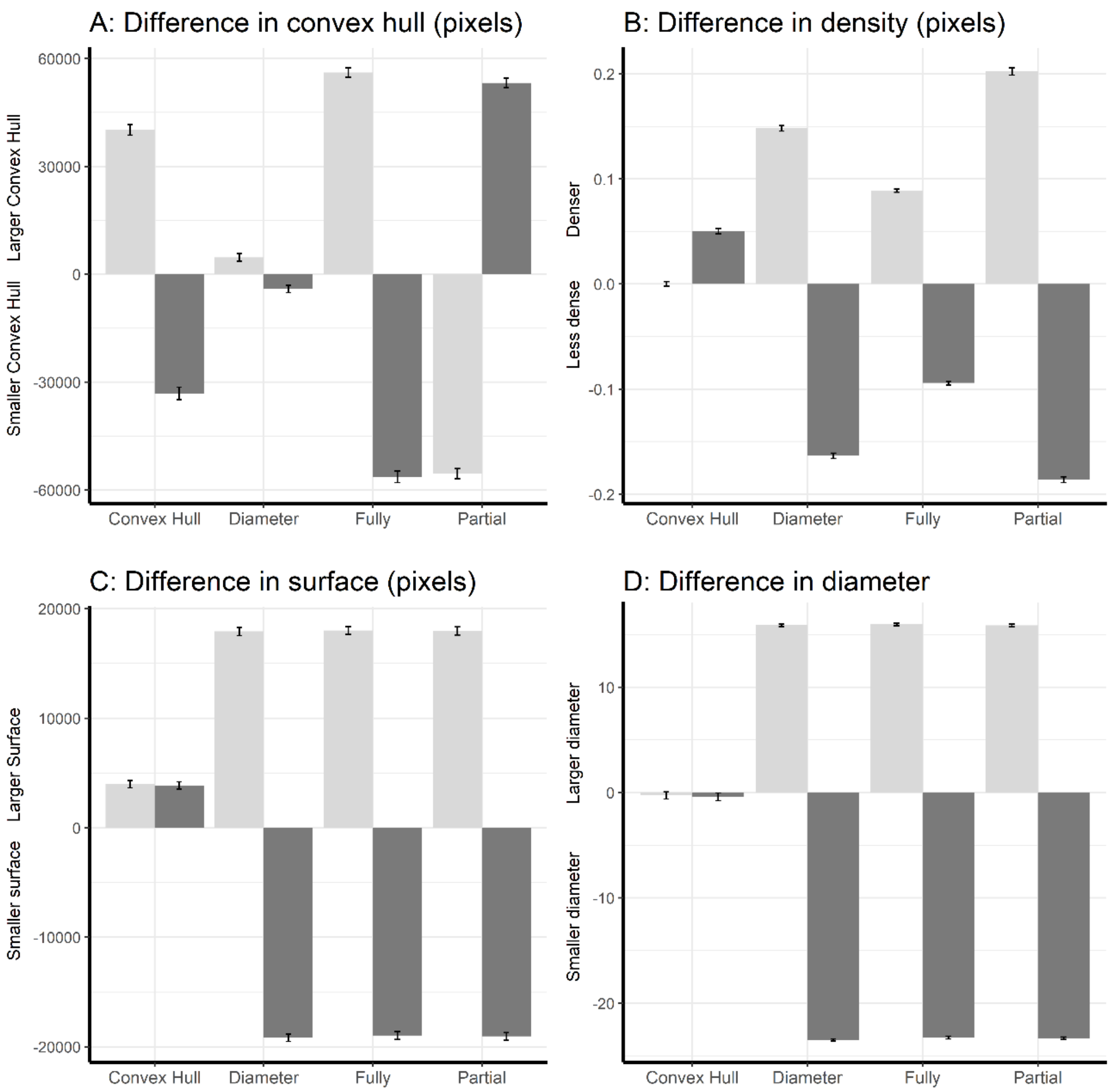


Figure 2: The difference in visual properties of all trials for each visual cue condition separately. Each panel depicts the difference in a certain visual property for the stimuli containing more dots relative to the stimuli containing less dots. The light and dark grey bars represent the difference in visual properties of each number pair for the congruent and incongruent trials respectively, calculated as the visual cue of the larger number minus the visual cue of the smaller number. Error bars are displaying standard error of the mean.

2.2.2. Dyscalculia Screener. The Dyscalculia Screener is a standardized test originally developed to assess mathematical abilities in children between the age of 6 and 14 (Butterworth, 2003). Standardized scores have been interpolated for adults and successfully used to assess mathematical proficiency in adult populations (Cappelletti et al., 2011, 2014; Cappelletti & Price, 2014). The test comprises four item-timed tasks which are divided into two subscales. The capacity subscale involves a dot enumeration task and a number comparison task, the achievement subscale contains two verification tasks (mathematical addition and multiplication). Every subtask starts with detailed instructions and several practice trials with feedback. For the actual test trials no feedback is provided. Based on participants' reaction time and accuracy, individual Stanine Scores and Standard Age Scores can be computed. In previous studies, adult participants with below average results on either of the tasks of the Capacity Subscale (Dot Enumeration and/or Numerical Stroop) have been classified as dyscalculic and this result has been confirmed by further diagnostic tools (Cappelletti et al., 2011, 2014; Cappelletti & Price, 2014). Since it is still unclear whether visual cues have the same or different effect on participants with low mathematical proficiency (as raised by Leibovich, Katzin, et al., 2016), we excluded participants who had a stanine score equal to or lower than 3 on either of the two tests of the Capacity Subscale.

2.3 Procedure

First, participants completed a dot comparison task with two different presentation modes. In both presentation modes the same dot arrays were shown to the participants, the only difference was whether trials of different types were presented in a blocked or mixed fashion. During blocked presentation mode, the four visual cue conditions were presented in separate blocks and their order was counterbalanced across participants. This part of the experiment is the replication of the original study by Gebuis & Reynvoet (2012a). During mixed presentation mode, trials of all four conditions randomly alternated. The order of the presentation mode (blocked and mixed) was also counterbalanced across participants. The experiment started with detailed instructions and six practice trials with feedback. After the practice trials no feedback was provided to the participants. Each trial began with a green fixation cross presented for 500 ms on the computer screen followed by a dot array for 300 ms, a blank screen for 500 ms and the second dot array for 300 ms. Then a red fixation cross appeared on the screen and remained visible until participants pressed one of the response buttons (Figure 3). They were instructed to press the left-CTRL button if the first array was more numerous and the right-CTRL button if the second dot array was more numerous. They were asked to respond as quickly and as accurately as possible. One presentation mode consisted of four blocks each with 192 trials. In total, each participant received six practice trials and 1536 experimental trials (192 trials per block × 4 blocks × 2 presentation modes). It should be noted that participants were not aware of the different presentation modes. From their point of view, they completed eight blocks of the same dot comparison task. Hence, instructions and practice trials were presented to them only once, at the beginning of the experiment. They had the opportunity to take a break after each block. After the dot comparison task they completed the Dyscalculia Screener. The duration of the experiment was 1.5-2 hours in total.

2.4 Statistical analysis

We calculated the percentage of correct responses for each presentation mode (blocked versus mixed), type of visual cue manipulation (1: convex hull, 2: diameter, 3: fully, 4: partially in/congruent) and congruency (congruent versus incongruent). In order to determine whether we could replicate the original results and whether there was a significant difference between the blocked and mixed presentation modes, accuracy data were subjected to a 3-way repeated measures ANOVA with factors presentation mode

(2), visual cue condition (4) and congruency (2). We report Greenhouse-Geisser corrected results, whenever the assumption of sphericity is violated.

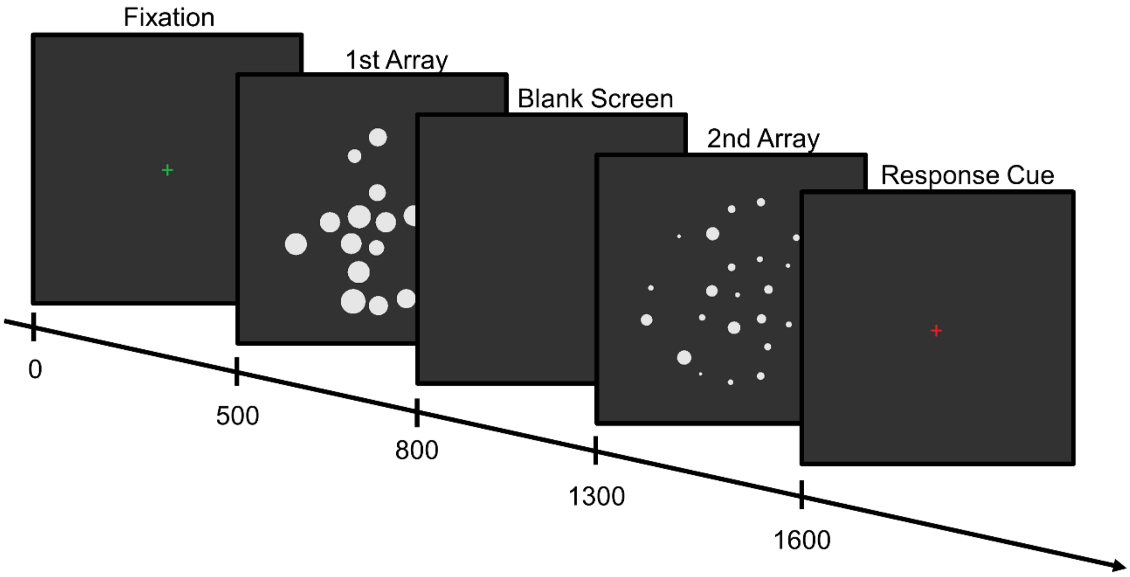


Figure 3: Experimental design with timing information.

3. Results

The analysis of variance revealed a main effect of presentation mode ($F(1,30)=4.69$, $p<.05$, $\eta_p^2=.14$), a main effect of visual cue condition ($F(1.81,54.29)=27.56$, $p<.001$, $\eta_p^2=.48$) and a main effect of congruency ($F(1,30)=21.25$, $p<.001$, $\eta_p^2=.42$). Main effects were qualified by a significant two-way interaction between visual cue condition and congruency ($F(1.66,49.80)=59.45$, $p<.001$, $\eta_p^2=.66$) and a three-way interaction between presentation mode, visual cue condition and congruency ($F(2.04,61.33)=6.99$, $p=.002$, $\eta_p^2=0.19$).

To follow-up the 3-way interaction, we conducted a 2-way analysis of variance with the factors visual cue conditions (4) and congruency (2), separately for the blocked and mixed presentation modes. Both ANOVAs revealed the same results: a main effect of visual cue condition (blocked: *F*(2.22,66.46)=21.01, *p*<.001, η_p^2=0.41, mixed: *F*(2.10,62.94)=21.24, *p*<.001, η_p^2=0.42), a main effect of congruency (blocked: *F*(1,30)=21.85, *p*<.001, η_p^2=0.42, mixed: *F*(1,30)=19.65, *p*<.001, η_p^2=0.40) and a significant 2-way interaction between visual cue condition and congruency (blocked: *F*(2.05,61.55)=43.68, *p*<.001, η_p^2=.59; mixed: *F*(1.51,45.18)=62.13, *p*<.001, η_p^2=.67) (Figure 4). Since the pattern of congruency effects is the same in both presentation modes (see Figure 4), the three-way interaction of presentation mode × visual cue conditions × congruency dominantly reflects a moderation in the congruency effects. Therefore, we report four two-way ANOVAs ([blocked vs. mixed] x [congruent vs. incongruent]) for every visual cue condition. We also report follow up t-tests for the single congruency effects (congruent vs. incongruent), separately for each presentation mode and visual cue condition.

3.1 Convex hull in/congruent condition

The 2-way analysis of variance conducted for the convex *hull in/congruent visual cue condition* revealed a significant main effect of congruency (*F*(1, 30)=88.52, *p*<.001, η_p^2=.75) and a significant interaction of presentation mode and congruency (*F*(1,30)=19.55, *p*<0.001, η_p^2=.40). The main effect of congruency was due to better performance on congruent trials (*M*=87.2, *SD*=1.26) than on incongruent trials (*M*=74.65, *SD*=1.72, see also Figure 4). Participants made significantly more correct responses when the array with more dots had a larger convex hull than when it had a smaller convex hull (all other visual cues equated). The interaction of presentation mode and congruency was due to a significantly larger congruency effect in the mixed presentation mode (congruent: *M*=88.71, *SD*=6.93, incongruent: *M*=72.45, *SD*=11.31) than in the blocked presentation mode (congruent: *M*=85.69, *SD*=8.29, incongruent: *M*=76.85, *SD*=9.71). The congruency effects ([congruent vs. incongruent]) were significant in both presentation modes (blocked: *t*(30)=5.87, *p*<.001, Cohen's *d*= 1.06, mixed: *t*(30)=9.89, *p*<.001, Cohen's *d*=1.78)

3.2 Diameter in/congruent condition

The ANOVA showed a significant main effect of presentation mode (*F*(1,30)=6.72, *p*<.05, η_p^2=.18), a significant main effect of congruency (*F*(1,30)=21.95, *p*<.001, η_p^2=.42) and a significant interaction of presentation mode and congruency (*F*(1,30)=7.52, *p*<.05, η_p^2=.20, see also Figure 4). The main effect of presentation mode was the result of better performance in the blocked (*M*=77.52, *SD*=1.68) than in the mixed presentation mode (*M*= 74.90, *SD*=1.73). The main effect of congruency was due to lower performance on congruent (*M*=64.80, *SD*=3.71) than on incongruent trials (*M*=87.62, *SD*=1.85). These results show that participants gave more correct responses when the array with more dot was associated with smaller visual cue (incongruent trials) which is in line with the results of the original study by Gebuis & Reynvoet (2012a). The interaction was again the result of a smaller congruency effect in the blocked (congruent: *M*=67.55, *SD*=3.67, incongruent: *M*=87.60 *SD*=1.90) than in the mixed presentation mode (congruent: *M*=62.16 *SD*=3.93, incongruent: *M*=87.63, *SD*=1.96). Follow-up t-tests showed that the congruency effects ([congruent vs. incongruent]) were significant during both presentation modes (blocked: *t*(30)=-4.22, *p*<.001, Cohen's *d*=-.76, mixed: *t*(30)=-4.95, *p*<.001, Cohen's *d*=-.89).

3.3 Fully in/congruent condition

The ANOVA did not yield any significant results (presentation mode: $F(1,30)=2.96$, $p=.10$, $\eta_p^2=.09$, congruency: F(1,30)=1.83, p=.19, $\eta_p^2=.06$, presentation mode × congruency: $F(1,30)=1.78$, $p=.19$, $\eta_p^2=.06$). T-tests on the congruency effects have not revealed any significant results either (blocked: $t(30)=-.87$, $p=.39$, Cohen's $d=-.16$, mixed: $t(30)=-1.6$, $p=.12$, Cohen's $d=-.29$).

3.4 Partially in/congruent condition

The analysis conducted for the *partially in/congruent condition* revealed a significant main effect of congruency ($F(1,30)=79.66$, $p<.001$, $\eta_p^2=.73$) and a significant interaction of presentation mode and congruency ($F(1,30)=4.75$, $p<.05$, $\eta_p^2=.14$). The main effect of congruency was a result of lower performance on *partially congruent trials* (diameter congruent, convex hull incongruent), (*M*=49.56, *SD*=4.15) than on *partially incongruent trials* (diameter incongruent, convex hull congruent), (*M*=92.00, *SD*=1.16). Again, the interaction of presentation mode and congruency was due to an increase in congruency effect from blocked (congruent: *M*=50.47, *SD*=4.14; incongruent: *M*=91.03, *SD*=1.42) to mixed presentation mode (congruent: *M*=48.66, *SD*=4.29; incongruent: *M*=92.77, *SD*=1.06). The congruency effects were significant in both presentation modes (blocked: $t(30)=-8.48$, $p<.001$, Cohen's $d=-1.52$, mixed: $t(30)=-4.99$, $p<.001$, Cohen's $d=-1.63$) Altogether, these findings indicate that participants gave more correct responses, when the array with more dots had a larger convex hull but smaller average diameter, aggregate surface and density *(partially incongruent trials)*. This pattern of results corresponds to the original findings by Gebuis & Reynvoet (2012a).

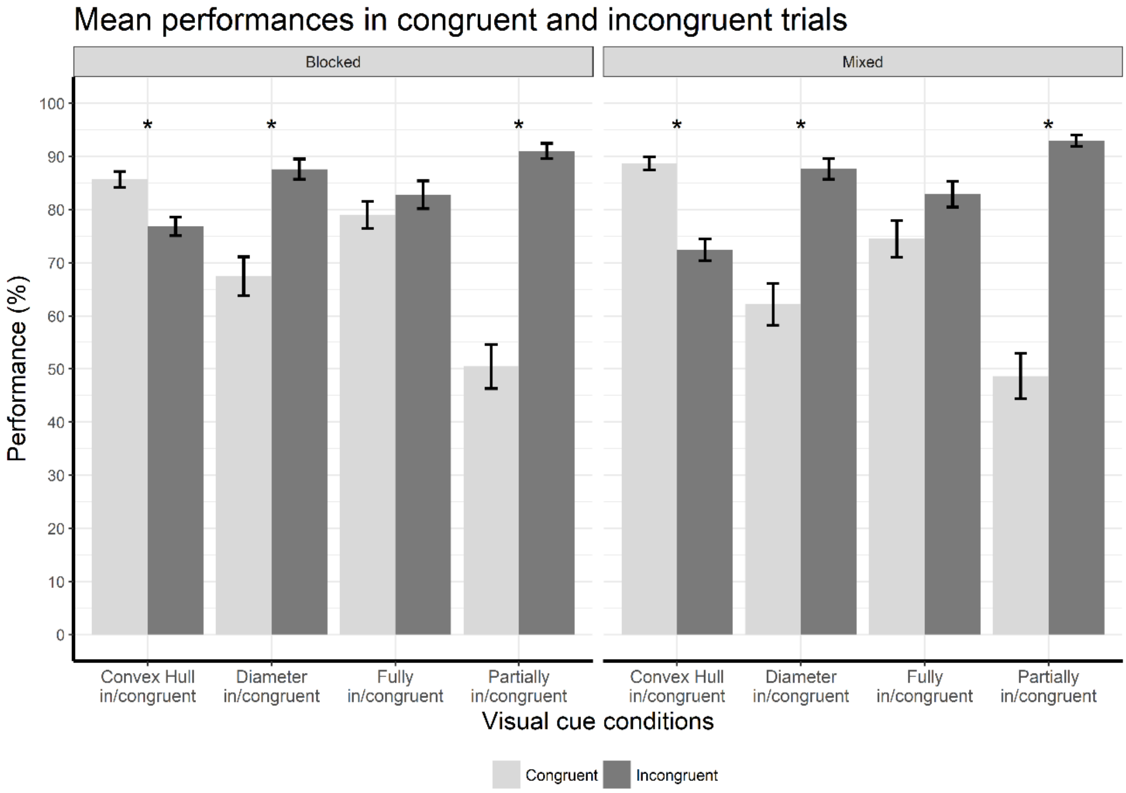


Figure 4: Performance (%) for each congruency and visual cue condition separately for the blocked and mixed presentation modes. In both presentation modes congruency effects were found in all conditions except the fully in/congruent condition. Light and dark grey bars represent congruent and incongruent trials, respectively. They also correspond to the light and dark grey bars in Figure 2. The asterisks represent significant results ($p_{adj} < .05$) Please note that *partially congruent trials* were congruent in diameter (as well as related visual cues) and incongruent in convex hull. Whereas, *partially incongruent trials* were incongruent in diameter (as well as related visual cues) and congruent in convex hull. Error bars are displaying standard error of the mean.

4. Discussion

The current study had two objectives. First, we examined whether changing the trial history by mixing trials of different types can alter the pattern of congruency effects shown by Gebuis & Reynvoet (2012a). Second, we were particularly interested in the reversed congruency effect in the diameter condition, when larger dots were estimated less numerous because this result is difficult to reconcile with findings from single-cell studies which would suggest the opposite (Nieder, 2016). We wanted to determine if the reversed diameter effect in the original study was merely a by-product of the block-wise presentation mode or whether it could be replicated when trials of different types are mixed together. For this purpose, we showed participants different trial types in a mixed fashion and compared it to the exact replication of the original study, where trials with different types of visual cue manipulation methods were presented in separate blocks.

We replicated the pattern of congruency effects found by Gebuis & Reynvoet (2012a). As in the original study, the same numerosities were used in all visual cue conditions, so any difference between congruent and incongruent trials can only be attributed to how the visual cues were manipulated within that condition. In the *convex hull in/congruent condition*, performance was better on congruent trials, i.e. when the more numerous dot array occupied a larger area (other visual cues were equated). We were also able to replicate the reversed congruency effect in the *diameter in/congruent condition*. Performance was better on incongruent trials, when the array with more dots had smaller average diameter, smaller aggregate surface and smaller density (convex hull was equated). When all visual cues were manipulated in the same direction in the *fully in/congruent condition*, no congruency effect was found. In the *partially in/congruent condition*, however, when the visual cues were manipulated in the opposite direction, the congruency effect increased. Gebuis & Reynvoet (2012a) draw the conclusion that when all visual cues are manipulated in the same direction *(fully in/congruent condition)*, the opposite effects of convex hull and diameter results in an attenuated congruency effect. However, when convex hull and dot diameter are manipulated in different directions *(partially in/congruent condition)*, an increase in the congruency effect is induced. Taken together, we were able to completely replicate these findings of Gebuis & Reynvoet

(2012a), including the reversed congruency effect in the *diameter in/congruent condition*. In contrast to our study, other studies using the same method to generate non-symbolic stimuli, included only dot arrays of the *fully* and the *partially in/congruent conditions* (Gilmore et al., 2013, 2016; Leibovich, Vogel, et al., 2016; Smets et al., 2015; Szűcs, Nobes, et al., 2013). Hence, our study is the first to reproduce the congruency effect in the *convex hull* and the reversed congruency effect in the *diameter in/congruent condition*.

We also compared performance and congruency effects between blocked and mixed presentation modes. As mentioned above, the pattern of congruency effects did not differ between them. However, congruency effects were greater when trials of different types were mixed together and this increase in congruency effect was significant in the *convex hull, diameter* and *partially in/congruent* conditions. It is important to note that these differences between the two presentation modes can only be attributed to how trials were presented to the participants (block-wise vs. mixed together), since the same dot arrays were shown during both tasks. The fact that the same pattern of congruency effects was found when trials of different types were mixed together supports the notion that convex hull and diameter have opposite effects on numerosity processing. It also shows that this is a stable effect and not simply a byproduct of the block-wise stimulus presentation method. Moreover, the increased congruency effects demonstrate that reliance on visual cues became larger when trials of different types were mixed together. Taken together, these results support that continuous magnitudes play an important role in numerosity processing and indicate that their role is not as straightforward as previously thought.

Specifically, it is noteworthy that whereas convex hull and numerosity are combined in a “greater is more” fashion, diameter and numerosity are combined in a “smaller is more” fashion. If the integration of numerosity and continuous magnitudes occurred on a basic neural level, one would expect a “greater is more” rule with all continuous magnitudes. It is possible that the reversed congruency effect in the diameter condition is a result of past experience and reflects a learning process about the relationship between numerical and non-numerical parameters. In real life, there tends to be a negative correlation between item diameter and numerosity when other sensory cues are equal. For example, when one of two equal-sized baskets is filled with apples and the other one with peas, then there must be in total more pieces of peas than apples. It is possible that participants have learned this negative relationship between the size of the items and their numerosity and applied what they have learned in the experiment when comparing dot patterns. This hypothesis is supported by the stability of the reversed diameter effect: it is not only present across different trial presentation modes (blocked vs. mixed) but also across different dot generation protocols. In fact, the effect has been replicated by studies using the Panamath method for generating dot comparison stimuli (Clayton et al., 2015; Norris et al., 2019). This is an important finding as some concerns have been raised about the replicability of findings in dot comparison tasks across different dot generation protocols (Inglis & Gilmore, 2014; Smets et al., 2015). There is only a single study in which larger dot diameter *increased* (rather than decreased) perceived numerosity in dot patterns (Salti et al., 2016). Yet, this contradiction can be resolved when one looks at the numerosity ranges used in this study. Whereas our study as well as all other studies finding a reversed diameter effect presented numerosities from the estimation range, Salti et al. (2016) used numerosities from the subitizing range

(2-4). Thus, it seems that within subitizing range dot diameter is integrated during numerosity processing the same way as other continuous magnitudes and this pattern is reversed in the estimation range. This differential effect of diameter on numerosity judgements further supports the notion that numerosities in the estimation and subitizing ranges are processed differently (Feigenson et al., 2004; Hyde & Spelke, 2009, 2011, 2012; Plodowski et al., 2003). A possible explanation is that diameter is combined with numerosity in a "greater is more" fashion in the subitizing range which is in line with findings of single-cell studies while the reversed congruency effect ("smaller is more") in the estimation range is possibly a result of past experience and reflects a learning process about the relationship between numerical and non-numerical parameters. This possibility, that learning is crucial in the integration of visual cues and numerosity, has been raised by several researchers (Gebuis et al., 2016; Leibovich, Katzin, et al., 2016; Mix et al., 2002).

Whereas the overall pattern of congruency effects was identical in both presentation modes, we could show that the influence of continuous magnitudes on performance increased when different trial types were presented in a mixed fashion. When a trial was preceded by trials of different types (as during mixed presentation mode) participants relied more on sensory cues than they did when the trial was preceded by trials of the same type (as during blocked presentation mode). In terms of Sensory Integration Theory this finding can be interpreted as an increase in weights given to the continuous magnitudes. Thus, our results add to the growing body of evidence suggesting that these weights are dynamic and applied in an adaptive and flexible way (Leibovich et al., 2015; Leibovich, Kallai, et al., 2016; Leibovich, Katzin, et al., 2016; Leibovich-Raveh et al., 2018). As mentioned before, different instructions, task difficulty, exposure duration and task context can alter the effect of continuous magnitudes on dot comparison performance. The differential effect of continuous magnitudes between blocked and mixed presentation modes in our study demonstrates that the weighing process depends not only on these factors but also on experiences participants have made with non-symbolic numerosities, including very recent ones in previous experimental trials.

Sensory Integration Theory assumes that the process of weighing the different continuous magnitudes when estimating the number of dots in a pattern is essentially a perceptual one. It is also possible, however, that changing the trial history has an impact on numerosity processing at some point later in the processing line. Leibovich et al. (Leibovich, Kallai, et al., 2016; Leibovich, Katzin, et al., 2016) emphasize the role of cognitive control abilities in numerosity processing: integration is necessary to allow us to use the natural correlation between continuous magnitudes and numerosity (e.g. congruent trials) but inhibition is required to suppress our bias to process the visual cues when the natural correlations are violated (e.g. incongruent trials). Thus, any detrimental influence of the visual cues on numerosity judgements may be a result of a deficient inhibition process. In line with this idea, studies found that both inhibition and integration play a role in numerosity processing (Cappelletti et al., 2015; Clayton & Gilmore, 2015; Fuhs & McNeil, 2013; Gilmore et al., 2013). An inhibition account of our results would suggest that it is more difficult to inhibit continuous magnitudes when different trial types are mixed together. The inhibition account may also provide an explanation for the reversed diameter effect: diameter is encoded the same way on the neuronal level as other continuous magnitudes but inhibition is required to overcome biases induced by dot size when making numerosity comparisons. The exact role of inhibition should be further investigated possibly by testing individuals of different age groups that have different levels of inhibitory control.

It should also be pointed out that the increased congruency effects during the mixed compared to the blocked presentation mode have practical implications. Dot comparison tasks are the most dominant methods to assess the acuity of the Approximate Number System. Although methods have been developed which aim to quantify and exclude the influence of visual cues on the ANS acuity measurement (DeWind et al., 2015), our study shows that not only the visual cues but also the presentation method may increase the reliance on sensory properties. This in turn might lead to an incorrect estimation of the ANS acuity. Our data show that this might especially be problematic when trials of different types are mixed together (e.g. Fazio, Bailey, Thompson, & Siegler, 2014; Gomez et al., 2015; Tokita & Ishiguchi, 2013). Consequently, the intention of reducing the reliance of continuous magnitude by mixing trials of different visual cue manipulation might have, in turn, led to the exact opposite – namely that the reliance on them has increased.

The role of continuous magnitudes in numerosity processing is a subject of debate in the current literature. Even though, very recent neurophysiological studies emphasize the superiority of numerosity over continuous magnitudes very early in the processing stream (Park et al., 2015), studies investigating the role of continuous magnitudes – including the present one – confirm repeatedly, that these indeed have a great impact on numerosity judgements. How great and complex this impact might be, can be best illustrated by performance in the partial condition. When visual cues follow a specific pattern, performance is close to a ceiling effect (i.e. *partially incongruent trials* = convex hull congruent, diameter incongruent) but when visual cues are combined in the opposite manner, surprisingly performance drops close to chance level (i.e. *partially congruent trials* = convex hull incongruent, diameter congruent). This pattern shows that visual cues have a massive effect on the perception of numerosities if they are manipulated in a certain manner. The fact that these same patterns of congruency effects were found during blocked and mixed presentation modes indicates that visual cues have a large and stable impact on numerosity judgments at some point in the processing line.

In sum, our results further support the notion that people integrate various continuous magnitudes when making numerosity judgements. Moreover, this integration process seems to be complex and adaptive. It seems to depend not only on the type of continuous magnitudes but also on experiences participants have made with non-symbolic numerosities, including very recent ones in previous experimental trials. It still needs to be investigated whether this process is essentially a perceptual one, that includes weighing of sensory cues, or whether higher-order processes, such as inhibition, are also involved.

Study 2

Electrophysiological correlates of the interaction of physical and numerical size in symbolic number processing: New insights from a hybrid go/nogo numerical Stroop task

Judit Pekár, Wiebke Hofmann, Balázs Knakker, Sascha Tamm & Annette Kinder

This PDF includes the submitted version of the manusciprt. Please access the original publication via the DOI provided below.

Pekár, J., Hofmann, W., Knakker, B., Tamm, S., & Kinder, A. (2023). Electrophysiological Correlates of the Interaction of Physical and Numerical Size in Symbolic Number Processing: Insights from a Novel Go/Nogo Numerical Stroop Task. Brain Sciences, 13(5), 702. https://doi.org/10.3390/brainsci13050702

1. Introduction

A central question in research on numerical cognition is how numerical information and continuous magnitudes, such as physical size, interact. Key to finding answers to this question is the numerical Stroop task in which participants are presented with pairs of Arabic digits and their task is to decide which number is larger based either on their numerical size or on their physical size (Henik & Tzelgov, 1982). According to these two dimensions the number pairs may be congruent (the numerically larger number is also physically larger) or they may be incongruent (the numerically larger number is presented in smaller physical size). Besides the well-known numerical distance effects – i.e. shorter reaction times for large compared to small numerical distances – studies using this paradigm have repeatedly reported facilitation and interference effects which show that not only task-relevant information but also task-irrelevant information is processed (Gebuis et al., 2010; Henik & Tzelgov, 1982; Huang et al., 2021; Kadosh et al., 2007; Kaufmann et al., 2005, 2008; Soltész et al., 2011; Szűcs et al., 2009; Szűcs & Soltész, 2007, 2012; Tang et al., 2006; Yao et al., 2015). Thus, the paradigm has been used to investigate the cognitive processes that are involved in the interaction of task-relevant and task-irrelevant stimulus features in general. Furthermore, it has also been used in the narrower context of numerical research in order to investigate at what stage of processing the interaction of physical and numerical information occurs and related to this, whether numbers and other magnitudes are processed by a common magnitude system or by separate systems.

Two competing hypotheses have been formulated that aim to explain the interaction of numerical and non-numerical information in the numerical Stroop task.

According to the *early interaction account*, numerical and physical dimensions of the stimuli are processed by common neural representations in the intraparietal sulcus and interact at an “early” processing stage before the appropriate motor response is selected, prepared and executed. As opposed, the *late interaction account* suggests that numerical and size information are processed parallel by different neural substrates and the conflict occurs at a later stage upon response initiation in the motor cortex (Gebuis et al., 2010; Kadosh et al., 2007; Schwarz & Heinze, 1998; Szűcs & Soltész, 2007). Thus, an early interaction would support the idea that numerical and size information is subserved by shared neural substrates, while a late interaction would indicate distinct neural substrates. To differentiate between these two opposing accounts, researchers search for the locus of congruency effects in the numerical Stroop task while measuring EEG and comparing certain ERP components, such as the P3 and the lateralized-readiness potential (Gebuis et al., 2010; Kadosh et al., 2007). The P3 component is a positive-going deflection which is usually observed over centro-parietal electrode sites 300-500 ms after stimulus presentation and it reflects stimulus evaluation and categorization processes (Donchin, 1981; Gebuis et al., 2010; Kutas et al., 1977; McCarthy & Donchin, 1981). Thus, congruency effects on the P3 indicate that the interaction of numerical and size information occurs at the stage of stimulus processing (early interaction). In contrast, the lateralized readiness potential reflects selective motor activation which is larger over electrode sites contralateral to the response hand. It is calculated by subtracting ipsilateral activity from contralateral activity. Congruency effects on the LRP suggest that the interaction occurs at the response level (late interaction).

Using these neural markers Gebuis et al. (2010) investigated whether the interaction of numerical information and continuous magnitudes occurs at the motor response stage or rather prior to that, during processes which lead up to stimulus evaluation and categorization. They examined the peak latency and the peak amplitude of the P3 component together with stimulus- and response locked LRPs. The P3 peak appeared later for trials with small compared to large numerical distances as well as for incongruent trials compared to congruent trials. The amplitude of the P3 component was larger for large numerical distances and for congruent trials. Furthermore, they found congruency effects on the stimulus-locked but not on the response-locked lateralized readiness potentials. Taken together, their results suggest that physical size interacted with numerical size before the end of stimulus evaluation and before the preparation or initiation of a response started. This pattern of findings is in line with the early interaction account and a general magnitude system. Kadosh et al. (2007) found the same pattern of congruency effects on the P3 amplitude as Gebuis et al. (2010). However, they (Kadosh et al., 2007) also reported late interaction effects on the response-locked LRPs which suggest that the processing of numerical and size information may under certain circumstances be subserved by distinct rather than shared neural substrates. Consequently, the matter of early versus late interaction of numerical and physical size has not been entirely settled yet. Nevertheless, it is clear that studies using the numerical Stroop paradigm have repeatedly confirmed that numerical and physical size interact and that the interaction may occur either at the perceptual or at the motor level.

Even though numerous studies have investigated the locus of the interaction between numerical and physical size information in the numerical Stroop task, to the best of our knowledge none of them investigated the exact nature of the psychological processes that underlie this interaction. Thus, the question arises: What kind of psychological processes are involved in the behavioural and electrophysiological effects that are observed in the numerical Stroop task? A process that may underlie the influence of physical size on Arabic number processing is inhibition. For example, it is possible that the in numerical Stroop task the physical size of the number pairs is inhibited in order to ensure responding to numerical size. In line with this idea, executive functions, especially inhibition has been reported as an important factor in numerical processing and development. Even though a large number of studies have been investigating the relationship between inhibition and numerical processing (Clayton & Gilmore, 2015; Fuhs & McNeil, 2013; Gilmore et al., 2013), a common feature of them is that they usually rely on behavioural methods (for review Cragg & Gilmore, 2014).

Inhibition, however, is not a unitary but rather a multifaceted construct and recent studies have been focusing on disentangling the electrophysiological signatures of different types of inhibition (Brydges et al., 2012, 2013; Groom & Cragg, 2015; Vuillier et al., 2016; Xie et al., 2017). Related to this, a major question in inhibition research is how different types of inhibition contribute to various behavioural and developmental profiles. One of the most common methods to investigate the different types of inhibitory processes is administering the go/nogo task while measuring EEG. In this task participants are presented with a stream of very simple stimuli (e.g. the letters M and W) and they are asked to respond with button press to frequent go trials (e.g. M) and refrain from responding to infrequent nogo trials (e.g. W). The electrophysiological correlates of inhibitory control that are measured in the go/nogo task are the N2 and P3 ERP components. The N2 ERP component is a fronto-central, stimulus-locked, negative going deflection occurring 200-350 msec post-stimulus that is larger on nogo than on go trials (Bruin et al., 2001; Bruin & Wijers, 2002; Falkenstein et al., 1999; Groom & Cragg, 2015; Jodo & Kayama, 1992). As opposed, the P3 component is a positive-going deflection often found between 300-500 msec after stimulus presentation but – as opposed to the previously mentioned categorization-related P3 component – the P3 in inhibition tasks has a more anterior topography that is referred to as the nogo anteriorisation (NGA). The nogo anteriorisation is thought to reflect activation in premotor areas, inferior frontal cortex and in the cingulate region which led to the conclusion that the component reflects response inhibition (Enriquez-Geppert et al., 2010; Fallgatter et al., 1997, 2000, 2001, 2002; Fallgatter & Strik, 1999; Huster et al., 2011; Smith et al., 2006, 2007, 2010; Strik et al., 1998; Tekok-Kilic et al., 2001). Because in the go/nogo task both the N2 and P3 components are enhanced on nogo compared to go trials, originally both components were considered as a marker of response inhibition.

Studies which focus on disentangling the electrophysiological signatures of different types of inhibition usually include the go/nogo and Flanker tasks, as well as their modified or hybrid versions (Brydges et al., 2012, 2013; Groom & Cragg, 2015; Xie et al., 2017). In Flanker tasks, for example, participants are presented with five arrows and their task is to respond to the central target arrow while the flanking arrows may be congruent or incongruent with the central target arrow (e.g. →→→→→ and →→←→→, respectively). In the hybrid version of this task the arrows are further manipulated in a way, that they either require a motor response (go trials – e.g. arrows point to left/right) or they require refraining from responding (nogo trials – e.g. arrows point up/down). Such a trial arrangement allows to separate two types of inhibition and their electrophysiological signatures: nogo trials require response inhibition while incongruent trials require resisting interference from the flanking arrows. Groom & Cragg (2015) implemented such a hybrid task and found that the N2 was enhanced on incongruent compared to congruent trials while the P3 was larger in trials that required response inhibition. Similar findings were reported by Xie et al. (2017) who found that interference inhibition was associated with larger N2 negativities while trials requiring response inhibition induced larger positivity on the P3. In sum, the findings support the notion that response inhibition induced by nogo trials and response conflict induced by incongruent trials have differential effects on the N2 and P3 ERP components.

At this point, it is important to note, that various terms have been used to characterize the different types of inhibitory processes, e.g. response conflict and response inhibition (Groom & Cragg, 2015), interference suppression and response inhibition (Brydges et al., 2012, 2013; Vuillier et al., 2016), stimulus interference control and response interference control (Jongen & Jonkman, 2008), interference inhibition and response inhibition (Xie et al., 2017). Similarly to Xie et al. (2017), we use the terms interference inhibition and response inhibition. Here, interference inhibition refers to resisting interference from irrelevant or misleading information, while response inhibition refers to stopping a prepotent response. For example, if inhibition is implicated in the numerical Stroop task, then analogically to the previously described early and late interaction accounts, it is possible that inhibition occurs at an "early stage" meaning that interference from irrelevant physical size is inhibited presumably at the stage of perceptual processing. Another possibility is that inhibition occurs at a "later stage" meaning that a prepotent response induced by physical size must be inhibited. Similar interpretations have been made by Soltész et al. (2011) who compared the electrophysiological correlates of the congruency effects in the numerical Stroop task between children and adults. They reported similar facilitation effects on the P3 component in both groups but found a larger interference effect in children than in adults. They concluded that, because children have less well-developed response control and executive functions than adults, the larger interference effects are due to less efficient inhibition of irrelevant physical size information. Even though this interpretation seems plausible, it is important to note, that it is based on modulations of the P3 component which – as described earlier – in this context reflects stimulus evaluation and categorization processes and not cognitive control functions per se. Thus, it was not possible to directly assess the involvement of inhibition in the interaction of physical and numerical size in that study.

As mentioned before, numerous studies have been investigating the relationship between numerical processing and inhibition. However, a common feature of them is that they usually rely on behavioural methods (Cragg & Gilmore, 2014). For example, inhibition is measured in one task and then numerical cognition in another then the relationship between the performance on these two kinds of tasks is measured. These methods deliver important insights on the relationship between inhibition and numerical abilities but they are limited in terms of investigating the contribution of the different types of inhibitions and to assess whether such inhibitory processes are directly linked to the well-documented interaction of physical and numerical size in numerical Stroop tasks. On the one hand, there are a handful of papers that implement the numerical Stroop task to investigate the electrophysiological correlates of inhibition, however their major intent does not lay at answering questions about numerical processing per se but rather at deciphering the developmental profiles or at elucidating the functional definition of certain neural markers (Bryce et al., 2011; Szűcs & Soltész, 2012). On the other hand, ERP studies that investigated the locus of interaction between physical and numerical size in the numerical Stroop task did not address whether inhibition underlies this interaction

(Gebuis et al., 2010; Huang et al., 2021; Kadosh et al., 2007; Soltész et al., 2011; Szűcs & Soltész, 2007; Yao et al., 2015).

Interestingly, even though cognitive control functions, especially inhibition, have been implicated in numerical processing and the N2 component has been associated with cognitive control, or more specifically, with suppressing interference from irrelevant stimulus-features, only a handful of papers investigated how the N2 component is modulated by congruency in the numerical Stroop task. For example, Huang et al. (2021) asked participants to judge number pairs based either on their physical size (Which number is physically larger? – Size Task) as well as based on their numerical size (Which number is numerically larger? – Number Task). Interestingly, in the Size Task they found a facilitation effect on the N2 component, i.e. less negative N2 amplitude on congruent compared to neutral and incongruent trials. In the Number Task, however, they did not report any congruency effects on the N2 component. Another study by Yao et al. (2015) investigated the effects of long-term abacus-based mental training on children's numerical processing. They found that two years later children who received the training showed congruency effects on the N2 as well as on the P3 components while children of the control group showed congruency effects only on the P3. The authors concluded that abacus-based mental training strengthens the relationship between number symbols and magnitude representation which in turn leads to faster and more automatic numerical processing. In sum, not only the number of studies that investigate the N2 in numerical Stroop are scarce but also their findings are inconclusive.

In the present study we aim to fill in the aforementioned gaps and explicitly test whether inhibition underlies the interaction of physical size and numerical size in the numerical Stroop task by examining related electrophysiological correlates. Because the

N2 and P3 components are reliably elicited and because they have been shown to reflect interference inhibition and response inhibition in go/nogo as well as in hybrid tasks, we also created a hybrid paradigm and combined go/nogo task with the numerical Stroop task. We presented participants with congruent, neutral and incongruent Arabic number pairs and instructed them to press a button if the number on the one side is numerically larger and refrain from responding if the number on the other side is numerically larger while we measured their EEG, accuracy and reaction times. Creating such a hybrid task has three advantages. One, including a go/nogo manipulation into the task makes it possible to obtain an unconfounded measure of response inhibition. The original numerical Stroop Task is a classical two-choice task where participants always have to implement one overt motor response. Therefore, inhibition of one response is always confounded with the initiation of the other. An unconfounded measure of response inhibition can be obtained when participants have to choose between withholding the motor response and implementing a certain motor response but not choosing between two possible motor responses. Two, combining the go/nogo paradigm with the numerical Stroop task allows to directly address the role of inhibition and differentiate between interference inhibition and response inhibition in the interaction of physical and numerical information while the inhibition-related N2 and P3 ERP components can be measured. Furthermore, it also provides insights about the processing stage at which inhibition may occur in numerical processing. Three, including not only congruent and incongruent but also neutral trials makes it possible to separately examine the interference and facilitation effects on the ERP components. This is not possible in the classical inhibition tasks, e.g. in the Flanker task, because those usually include only congruent and incongruent trials.

In order to assess whether implementing this novel hybrid paradigm allowed to elicit the inhibition-related N2 and P3 components, first we contrasted go and nogo trials of large numerical distances (small numerical distances were always go trials, for more details and reasoning see Methods section). In accordance with the classical go/nogo tasks, we expected to observe larger negativities on the N2 component over fronto-central electrode sites for nogo compared to go trials and we expected larger positivities on the P3 components over centro-parietal electrodes for nogo compared to go trials. Second, we also analysed congruency effects on these components to investigate how physical size manipulations modulate the N2 and/or P3 components in this task. Third, in order to assess whether the repeatedly reported interference and facilitation effects of the numerical Stroop task can also be observed in this novel hybrid paradigm, we performed a mass univariate analysis to determine possible onset and offset latencies to assess at what point in time incongruent trials differ from neutral trials (interference) and congruent trials differ from neutral trials (facilitation). Additionally, we also investigated whether we could extend the findings by Gebuis et al. (2010) and Kadosh et al. (2007) with regards to the categorization-related P3 peak latency and amplitude and find the numerical distance and congruency effects on the categorization-related P3 component. To this end we compared peak latency and peak amplitude between small and large numerical distances and different congruency conditions on go trials and expected equivalent results to those by Gebuis et al. (2010) and Kadosh et al. (2007).

2. Methods

2.1 Participants

In total 23 individuals were included in the study (ten females, age: *M*=24.91, *SD*=3.93, range: 18-34 years) from which three were excluded due to technical problems occurring during data acquisition. All of the individuals were right-handed, had normal or corrected-to-normal vision and did not report any psychiatric or neurological disorder. They gave written informed consent and received either course credit or monetary compensation for the participation. The experimental procedure was approved by the local ethics committee of the Free University Berlin. Data from four participants were not included in the data analysis because more than 25% of their EEG data contained artefacts. Thus, the final dataset consisted of data from 16 participants (seven females, age: *M*=23.81, *SD*=3.40, range: 18-34 years).

2.2 Apparatus, stimuli and procedure

Trials were presented on a 23-inch monitor using PsychoPy (Peirce, 2007, 2009). Participants were presented with pairs of Arabic digits ranging from 1 to 9 and their task was to indicate whether the number on the left side or the number on the right side was numerically larger. We manipulated (1) the response type in order to create go and nogo trials, (2) the physical size of the numbers in order to create different congruency conditions, as well as (3) the numerical distance between the numbers in order to include small and large numerical distances. In terms of response type manipulation participants either had to press a button if the number on the one side was numerically larger (go trials) or they actively had to refrain from responding if the number on the other side was numerically larger (nogo trials, also see Figure 1B). In the entire experiment the ratio of go versus nogo trials was .75 and .25, respectively. All participants were instructed to press the right CTRL button with their right index finger on go trials. The side-to-response assignment was counterbalanced across the participants. As a result, the final dataset included 9 participants who responded with button press when the number on the left side was larger (no response otherwise) and 7 participants who had to push a button when the number on the right side was numerically larger (no response otherwise). Congruency conditions were created by manipulating the physical size of the numbers and presenting them in various font sizes (Figure 1B): on congruent trials the numerically larger number had a larger font size than the numerically smaller number (e.g. 7 2). On incongruent trials the numerically larger number was presented in smaller font size (e.g. 7 2) whereas neutral trials both numbers were presented in the same font size (e.g. 7 2). For the numerical distance manipulation, we created all available number pairs with a distance of one for the small distance condition (e.g. 1 2, 2 3, 3 4 and so on) and with a distance of five for the large distance condition (e.g. 1 6, 2 7, 3 8 and so on). These number pairs constituted the trials of interest and were included in the behavioural and EEG data analysis. For the analysis of go versus nogo, only large distance trials were included. These trials occurred equally often throughout the experiment. The number of congruent, neutral and incongruent trials was also equal both during the entire experiment as well as in the analysis. We also included filler trials in the task in order to (1) avoid expectation effects (e.g. trials with a distance of five are always nogo trials) and (2) to keep the ratio of go and nogo trials at .75/.25 throughout the experiment while also keeping number of trials that were included in the analysis equal. Filler trials could have all possible distances between 1 and 9 and they could be all possible combinations of go/nogo and congruency manipulations (Figure 1C). The viewing distance was about 65 cm. The stimuli were presented with a width of 0.8 and a height of 1.15 in degrees of visual angle for small, 1.32 and 1.94 degrees for large as well as with 0.97 and 1.5 degrees for neutral trials.

The experiment started with detailed instructions and eight practice trials with feedback. After the practice trials no feedback was given to the participants. Each trial began with a fixation cross presented for 1250-1500 msec with a stimulus onset asynchrony of at least 10 msec. This was followed by a number pair displayed for 200 msec and a fixation cross. The fixation cross was shown for a maximum of 1250 msec or until the participant pressed a button. In either case, a blank screen was presented for 300 msec before the next trial started (Figure 1A).

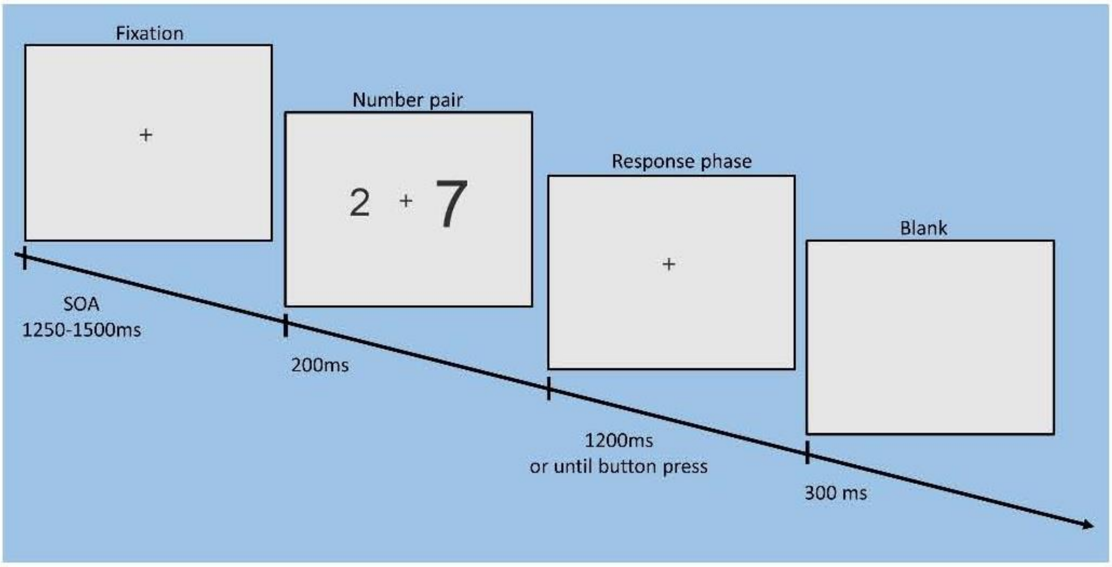
Participants completed six experimental blocks each consisting of 240 trials (1440 trials in total). Each block contained 144 trials of interest and 96 filler trials. The ratio of go/nogo trials was .75/.25. Congruent, neutral and incongruent trials were presented equally often. This arrangement resulted in 96 trials of interest for each response type (go versus nogo), for each distance (small versus large) as well as for each congruency condition (congruent, neutral, incongruent) (Figure 1C).

2.3 Recording and preprocessing of electrophysiological data

Electrophysiological data was recorded from 64 active electrodes placed according to the extended international 10-10 system (actiCAP system, BrainProducts, Munich). All electrodes were online referenced to the FCz while AFz served as ground electrode. Impedances were kept below 10 kΩ for reference and ground electrodes and below 20 kΩ for the active electrodes. The recordings were amplified using the BrainAmp system (BrainProducts, Munich). The sampling rate was 500 Hz.

EEG data was preprocessed using EEGlab (Delorme & Makeig, 2004). The data was filtered using a low-pass filter of 20 Hz, a high-pass filter of 0.1 Hz and a notch filter

A



B

go/nogo Manipulation	Numerical Stroop Manipulation		
	congruent	neutral	incongruent
Go trials (75%)	2 7	2 7	2 7
	Press if the number on the right side is larger		
Nogo trials (25%)	7 2	7 2	7 2
	Don't press if the number on the left side is larger		

C

Number of trials	Trials of interest (analysis)									
	go						nogo			9*96= **864**
	small distance			large distance			large distance			
	cong	neut	incong	cong	neut	incong	cong	neut	incong	
	96	96	96	96	96	96	96	96	96	
	Filler trials (with varying distances)									
	go						nogo			3*168+3*24= **576**
	cong		neut		incong		cong	neut	incong	
	168		168		168		24	24	24	
	6*96+3*168= **1080 (75%)**						3*96+3*24= **360 (25%)**			TOTAL # trials **1440**

Figure 1: a) Experimental design with timing information. b) Example trials for each response type and congruency manipulation. c) Number of trials by trial types in the experiment. The trials of interest had always a distance of one or five, while filler trials could have all distances between one and nine.

of 50 Hz (each 24 dB/oct). Technical artefacts were removed manually and bad channels were interpolated using spherical lines before carrying out independent component analysis to remove ocular artefacts from the EEG signal. The EEG signals were rereferenced to the average of all included electrodes and then segmented into epochs from 200 msec prior to 800 msec after the stimulus presentation. The 200 msec pre-stimulus interval was used for the baseline correction. Trials with artefacts and filler trials were removed from the data analysis. Artefactual trials were detected using EEGlab's moving window peak-to-peak method with a window width of 200 msec, a window jump of 50 msec and a threshold of 75 µV. Average ERPs were conducted for each participant, trial type and electrode.

Based on previous literature that investigated the N2 and P3 components in (hybrid) go/nogo paradigms, we expected effects on midline electrodes (Bruin et al., 2001; Bruin & Wijers, 2002; Brydges et al., 2012, 2013; Groom & Cragg, 2015). Visual inspection confirmed our expectations, thus the N2 event-related potential was measured as the maximum negative amplitude between 250-350 msec on electrodes Fz, FCz and Cz. The inhibition-related P3 component was defined as the maximum positive amplitude between 400-600 msec on fronto-central and centro-parietal electrode sites Fz, FCz, Cz and CPz. We also measured the categorization-related P3 component which was defined according to Gebuis et al. (2010) and was calculated as the peak amplitude and the peak latency between 300-800 msec on electrode Pz.

3. Results

3.1 Behavioural Results

3.1.1 Reaction time – Distance effects on go trials: The 2-way repeated measures ANOVA showed a main effect of distance (*F*(1,15)=154.67, p<0.001, η_p^2=0.91) and a main effect of congruency (*F*(2,14)=190.05, p<0.001, η_p^2=0.96) (Figure 2). The main effect of distance was a result of faster responses for large (*M*=427.51, *SD*=13.33) compared to small numerical distances (*M*=467.19, *SD*=13.74). Follow-up tests on the main effect of congruency revealed shorter reaction times for congruent than for neutral (*t*(15)=-4.30, *p*<.001, *Cohen's d*=-1.08, congruent: *M*=426.69, *SD*=13.13, neutral: *M*=437.02, *SD*=13.65) and shorter reaction times for congruent than for incongruent trials (*t*(15)=-19.31, *p*<.001, *Cohen's d*=-4.83, incongruent: *M*=478.33, *SD*=13.78) as well as shorter reaction times for neutral than for incongruent trials (*t*(15)=-16.02, *p*<.001, *Cohen's d*=-4.01).

3.1.2 Accuracy Data – Go/nogo effects: The 2-way analysis of variance on accuracy data have revealed a main effect of go/nogo (*F*(1,15)=22.57, p<0.001, η_p^2=0.60), a main effect of congruency (*F*(1.21,18.15)=24.10, p<0.001, η_p^2=0.62) and an interaction of go/nogo and congruency (*F*(1.17,17.59)=22.89, p<0.001, η_p^2=0.60) (Figure 3). The interaction was followed-up by two one-way ANOVAs, conducted on factor congruency separately for the go and nogo trials. They revealed that congruency effect was absent for the go trials (*F*(2,30)=1.0, p>0.05, η_p^2=0.06) but present for the nogo trials (*F*(2,30)=23.71, p<0.001, η_p^2=0.61). Pair-wise tests on the main effect of congruency for nogo trials showed that performance on incongruent trials (*M*=87.89, *SD*=9.19) was

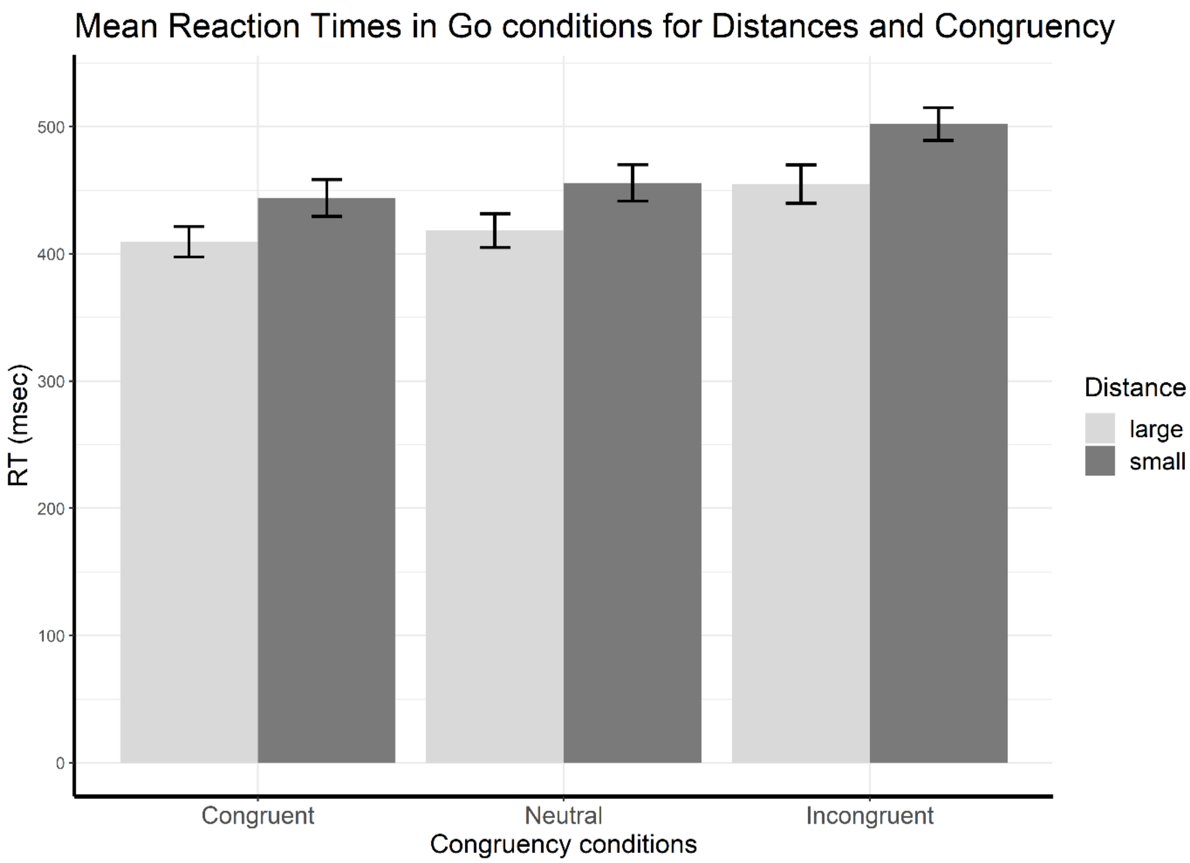


Figure 2: Reaction time results – mean reaction times on go trials for each distance and congruency condition. Error bars indicate standard error of the mean.

significantly lower than on congruent ($t(15)=4.81$, $p<0.001$, *Cohen's d*=1.20, $M=96.22$, $SD=4.04$) and neutral trials ($t(15)=5.51$, $p<0.001$, *Cohen's d*=1.38, $M=95.31$, $SD=5.55$), whereas the latter two did not differ from each other ($t(15)=1.33$, $p>0.05$, *Cohen's d*=.33). The main effect of go/nogo was due to significantly higher accuracy on go trials ($M=99.78$, $SD=.36$) than on nogo trials ($M=93.14$, $SD=5.87$) and follow-up t-tests showed that this effect was present for all congruency conditions (congruent: $t(15)=3.90$, $p<0.05$, *Cohen's d*=.97, neutral: $t(15)=3.40$, $p<0.05$, *Cohen's d*=.85, incongruent: $t(15)=5.24$, $p<0.001$, *Cohen's d*=1.31).

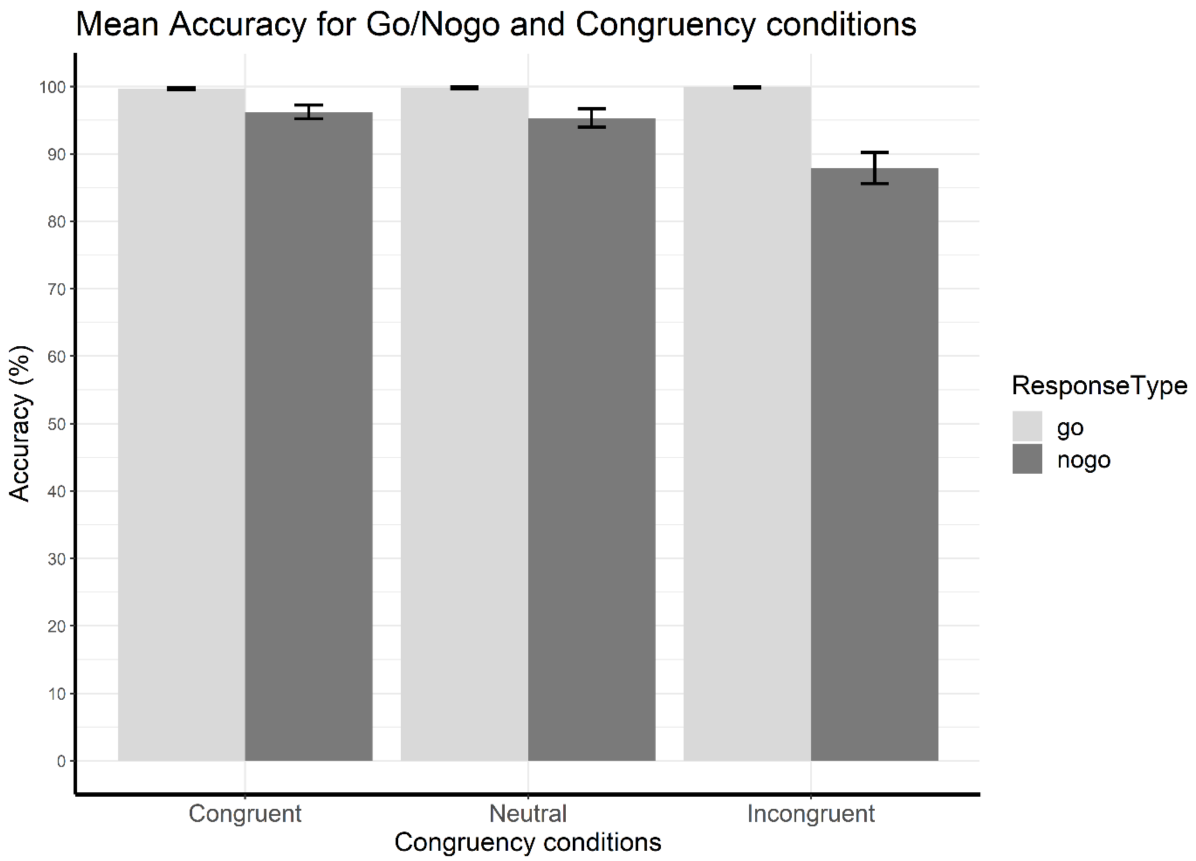


Figure 3: Accuracy results – Mean Accuracy for each go/nogo and congruency conditions. Error bars indicate standard error of the mean.

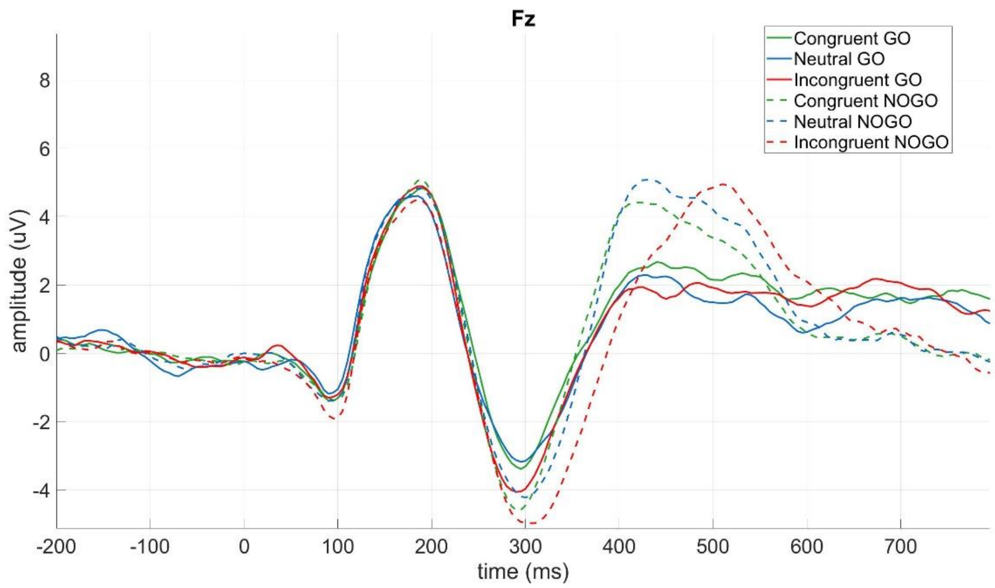
3.2 Event-related potentials: Peak Analysis

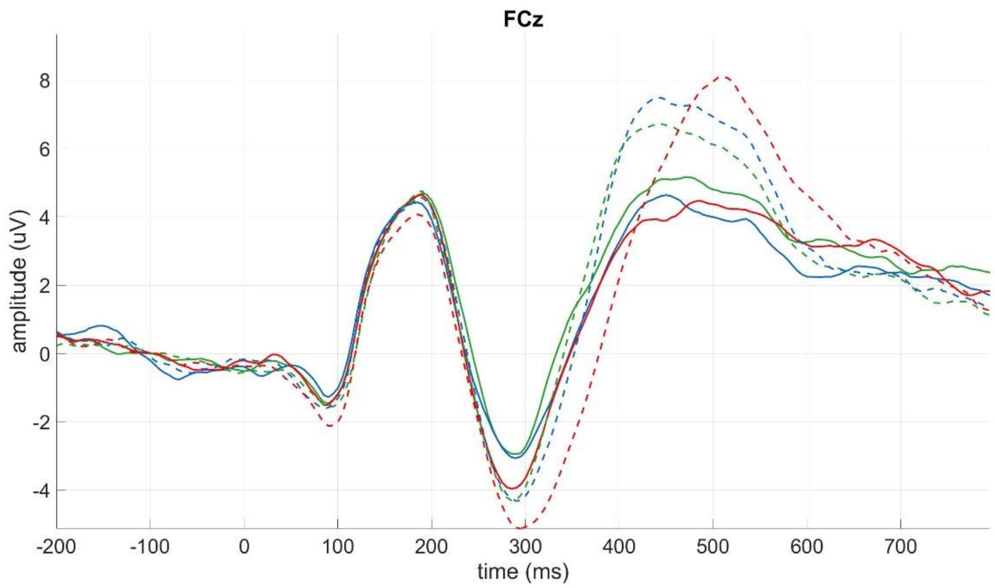
3.2.1 Modulation of the inhibition-related N2 by trial type

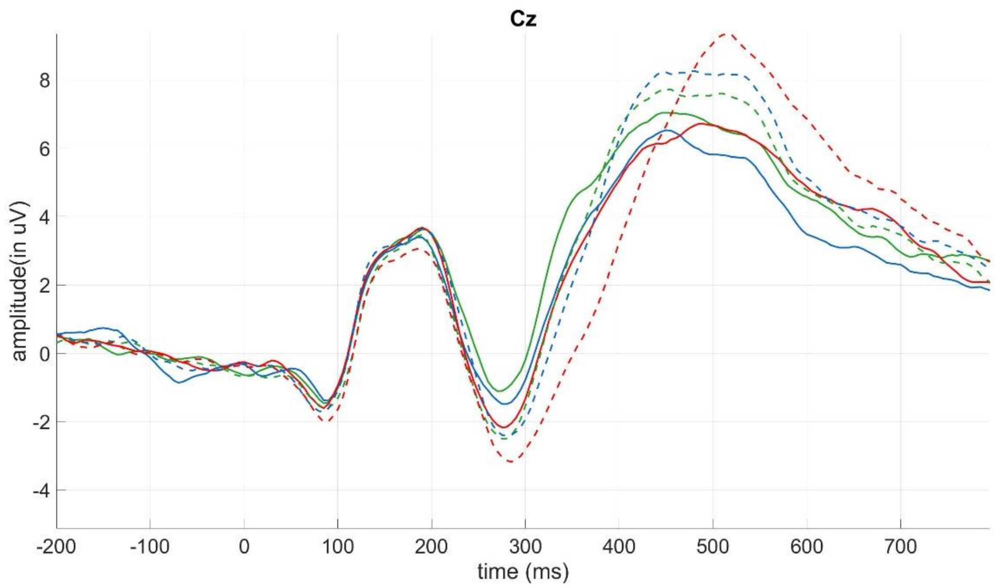
The 3-way ANOVA revealed a main effect of go/nogo (F(1,15)=10.47, p<0.05, η_p^2=0.41), congruency (F(2,30)=6.13, p<0.05, η_p^2=0.29) and electrode (F(2,30)=24.42, p<0.001, η_p^2=0.62) (Figure 4). None of the interactions were significant. The main effect of go/nogo was due to larger negative amplitudes on nogo than on go trials (M=-5.25, SD=3.89 and M=-3.69, SD=2.91, respectively) whereas the congruency main effect reflected significantly larger negative peak amplitudes on incongruent trials than on congruent (*t*(15)=3.12, *p*<.05, *Cohen's d*=.78, incongruent: *M*=-5.07, *SD*=3.37, congruent: *M*=-4.12, *SD*=3.32) as well as on neutral trials (*t*(15)=3.12, *p*<.05, *Cohen's d*=.78, neutral: *M*=-4.23, *SD*=3.41). Congruent and neutral trials did not differ from one another (*t*(15)=.32, *p*>.05, *Cohen's d*=.08). The main effect of electrode was caused by significantly less negative peak amplitudes on electrode Cz than on electrodes Fz (*t*(15)=-5.69, *p*<.001, *Cohen's d*=-1.42, Cz: *M*=-3.03, *SD*=3:12, Fz: *M*=-5.24, *SD*=3.22) and FCz (*t*(15)=-5.87, *p*<.001, *Cohen's d*=-1.47, FCz: *M*=-5.15, *SD*=3.82). Electrodes Fz and FCz did not differ from each other (*t*(15)=.79, *p*>.05, *Cohen's d*=-.07).

3.2.2 Modulation of the inhibition-related P3 by trial type

The 3-way ANOVA resulted in a main effect of go/nogo (*F*(1,15)=22.03, *p*<0.001, η_p^2=0.60), and a main effect of electrode (*F*(1.53,22.89)=31.13, *p*<0.001, η_p^2=0.68) (Figure 4). The main effect of congruency did not reach significance (*F*(2,30)=.53, *p*>0.05, η_p^2=0.04). The main effect of go/nogo was due to significantly larger positive amplitudes on nogo trials than on go trials (*M*=8.68, *SD*=2.74 and *M*=6.56, *SD*=3.48, respectively). Post hoc test on the electrode main effect revealed that peaks were significantly more positive on electrodes Cz (*M*=8.99, *SD*=3.50) than on Fz (*M*=5.0, *SD*=2.92, *t*(15)=-6.55, *p*<.001, *Cohen's d*=-1.64) and FCz (*M*=7.71, *SD*=3.29, *t*(15)=-5.34, *p*<.001, *Cohen's d*=-1.3334) and also on CPz (*M*=8.79, *SD*=3.09) again compared to Fz and FCz (*t*(15)=-5.8884, *p*<.001, *Cohen's d*=-1.46 and *t*(15)=-2.64, *p*<.001, *Cohen's d*=-.66, respectively). These latter two also differed significantly with peaks significantly more positive on FCz than on Fz (*t*(15)=-5.87, *p*<.001, *Cohen's d*=-1.47).


Fz
Congruent GO
Neutral GO
Incongruent GO
Congruent NOGO
Neutral NOGO
Incongruent NOGO
amplitude (uV)
time (ms)


FCz
amplitude (uV)
time (ms)



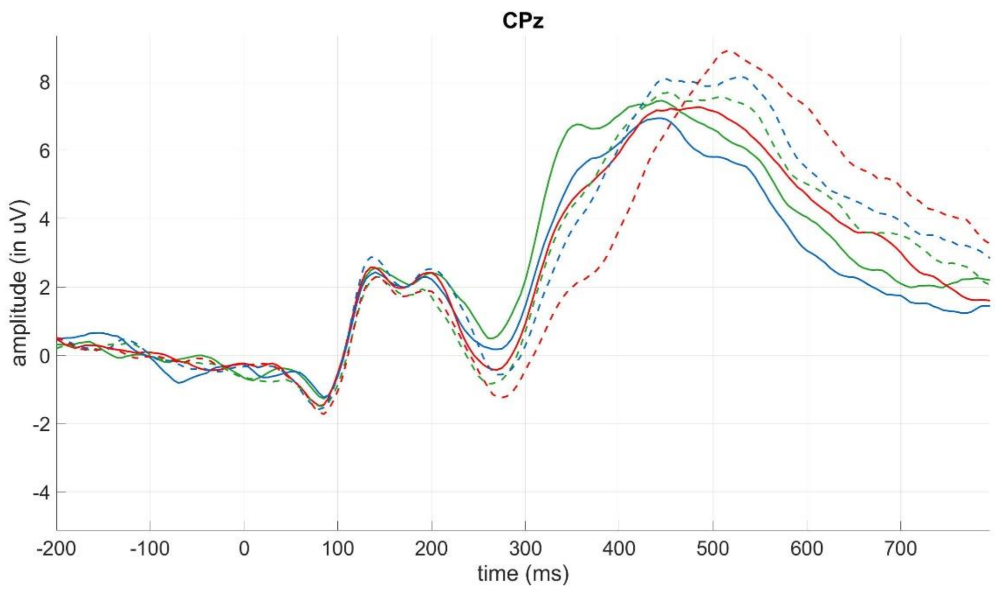


Figure 4: Go/nogo and congruency effects on the peak amplitude of the inhibition-related N2 and P3 components on electrodes Fz, FCz,, Cz and CPz. Positive up.

3.2.3 Modulation of the categorization-related P3 by trial type

The 2-way analysis of variance on the P3 peak latency (go trials) revealed a significant main effect of distance (*F*(1,15)=11.43, *p*<.05, η_p^2=0.43) and congruency (*F*(2,30)=6.57, *p*<.05, η_p^2=0.30) (Figure 5). The main effect of distance was a result of the peak appearing later for small (*M*=452.60, *SD*=92.67) than for large numerical distances (*M*=421.98, *SD*=66.44). The main effect of congruency was due to peaks appearing earlier for congruent (*M*=411.88, *SD*=64.09) than for neutral (*M*=435.47, *SD*=96.37) and for incongruent trials (*M*=464.53, *SD*=92.28). Post hoc analysis revealed that this difference was significant between congruent and incongruent trials (*t*(15)=-3.17, *p*<.05, *Cohen's d*=-.79). The 2-way analysis of variance on peak amplitude did not reveal any significant differences.

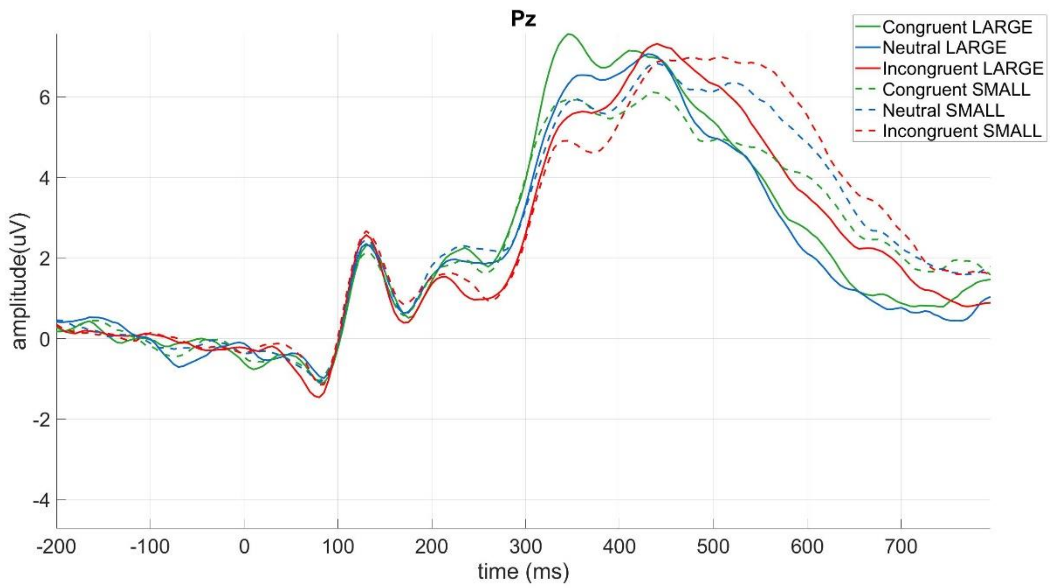


Figure 5: Effect of numerical distance and congruency on the peak latency of the categorization-related P3 component on electrode Pz

3.3 Event-related potentials: Mass Univariate Analysis

Permutation-based test analysis on difference waves in the go condition showed that congruent and neutral trials started to show differences on electrode CPz from 325 ms to 350 msec (critical t-scores: +/-4.25, that corresponds to a test-wise alpha level of 0.000706, total number of comparisons 484, total number of permutations 2500). No significant differences were detected between incongruent and neutral trials.

As opposed, on nogo trials significant differences were detected between incongruent and neutral trials starting at 360 msec on Fz electrodes which then spread to all other midline electrodes and lasted up to 420 msec on Cz and CPz (critical t-scores: +/-4.16, that corresponds to test-wise alpha level of 0.000832). On Nogo trials, however, no significant differences between neutral and congruent trials were detected.

4. Discussion

The aim of the present study was to investigate the role of inhibition in the interaction of numerical and physical size in the numerical Stroop task by examining related electrophysiological components. To this end, we introduced a novel hybrid paradigm and combined the numerical Stroop task and the go/nogo task. We presented participants with congruent, neutral and incongruent Arabic number pairs and asked them to press a button if the number on the one side of the screen was numerically larger and to refrain from responding if the number on the other side of the screen was numerically larger (Bruin et al., 2001; Bruin & Wijers, 2002; Henik & Tzelgov, 1982). This arrangement allowed us to measure the well-established inhibition-related N2 and P3 ERP components. Analysis of reaction time and accuracy data confirmed the standard behavioural effects reported by previous studies that implemented the numerical Stroop task or the go/nogo task. Participants responded faster and made less errors on large compared to small numerical distances, as well as on congruent compared to neutral and compared to incongruent trials. As for the go/nogo manipulation, participants made generally less errors on go than on nogo trials and in the case of nogo trials a congruency effect was also observed. Less errors were made on congruent, than on neutral and on incongruent trials. Beyond replicating basic behavioural effects, we had four objectives. First, we were interested whether this hybrid paradigm is suitable to elicit the classical go/nogo effects on the N2 and P3 ERP components. Second, by analysing the congruency effects on the inhibition-related N2 and P3 components, we wanted to investigate how physical size manipulations modulate the N2 and P3 components in this task. Third, in order to assess whether the repeatedly reported facilitation and interference effects of the numerical Stroop task can be observed in this hybrid paradigm, we implemented a mass univariate analysis and contrasted neutral trials with congruent and incongruent ones separately on go and nogo trials. And finally, we also examined whether we could extend the findings of Gebuis et al. (2010) and Cohen Kadosh et al. (2007) who reported congruency and distance effects on the amplitude and the peak latency of the categorization-related P3 component.

When looking at the effects of the go/nogo manipulation on the N2 and P3 ERP components, we found that nogo trials elicited more negative N2 as well as more positive P3 components than go trials. These electrophysiological findings are in accord with those observed in classical go/nogo tasks and hybrid tasks (Donkers & Van Boxtel, 2004; Enriquez-Geppert et al., 2010; Falkenstein et al., 1999; Jodo & Kayama, 1992; Nieuwenhuis et al., 2003, 2004; Randall & Smith, 2011; Smith et al., 2010). Thus, this study is the first one to show this effect on the N2 and P3 components in a hybrid go/nogo numerical Stroop task. Furthermore, it shows that this novel hybrid paradigm is suitable to elicit these inhibition-related ERP components and therefore to investigate whether and how inhibitory processes underlie the interaction of physical and numerical size. As opposed to classical inhibition tasks which can be described as simple decision tasks (e.g. respond when letter M presented but don't respond when the letter W is presented), in the current paradigm numerical judgements were made as participants had to decide which of the two presented numbers was numerically larger. Thus, the reported differences between go and nogo trials on the N2 and P3 components also show that implementing certain modifications to the go/nogo task is an effective tool to investigate inhibition in more complex cognitive processes as well.

When analysing the interaction of physical and numerical size on go and nogo trials, we found that congruency effects occurred on the N2 but not on the P3 component. Incongruent trials were more negative than congruent and neutral ones, whereas congruent and neutral trials did not differ from one another. Since we measured the N2 component as the peak amplitude between 250-350 ms, this pattern of results implicates that both physical and numerical size, as well as their relation to each another has to be processed prior to this peak. Furthermore, enhancement on the N2 amplitude was apparent only on incongruent trials, when the irrelevant stimulus dimension (physical size) was misleading and interfered with the numerical size. It was not present on neutral and congruent trials, when physical size did not contradict numerical size. As mentioned before, only a handful of papers investigated the N2 component in numerical Stroop task and their findings were inconclusive. Huang et al., (2021) did not report any effects on the anterior N2 with adult participants. Yao et al., (2015) found marginally more negative

N2 on frontocentral electrodes on congruent than on incongruent and neutral trials in children after receiving mental-abacus training but did not report any effects in the control group. Furthermore, the effects on the N2 component were also observed later in time than, around 370-470 ms, than in the inhibition tasks. As opposed to Yao et al. (2015) and Huang et al. (2021), using the hybrid go/nogo numerical Stroop paradigm, we found a congruency effect on the fronto-central N2 component in adults participants. Furthermore, the timing and the pattern of the N2 component in the current study are in accord with those reported by hybrid Flanker go/nogo tasks. Namely, that incongruent trials elicit more negative anterior N2 component than congruent trials (those tasks did not include neutral trials). Thus, it seems that when the irrelevant stimulus dimension induces interference, as in the case of incongruent trials, then this interference is inhibited between 250-350 ms. As opposed, when the irrelevant stimulus dimension does not induce interference, as on congruent and neutral trials, no evidence of interference inhibition was observed. At the same time, we also found no evidence for an interaction between physical size and numerical size on the P3 component as it was not modulated by congruency effects. Taken together, this pattern of results seems to support the early interaction account, namely that physical and numerical size interact at the stage of perceptual processing. Furthermore, it also implicates that the interference caused by misleading information of physical size on incongruent trials is probably resolved by inhibitory processes that occur very early, between 250-350 ms. Introducing the go/nogo condition was necessary to reliably elicit the inhibition-related N2 component and to obtain an unconfounded measure of inhibition, i.e. when the inhibition of an overt motor response is not confounded by the initiation of another, like in the classical numerical Stroop task. It is however, still unclear, how numerical distance affects the N2 component.

Even though both Yao et al. (2015) and Huang et al. (2021)included more than one distance, they did not investigate how distance may affect the N2 component. In the current study we included only number pairs with a numerical distance of five.

Concurrently manipulating numerical and physical size in this task made it possible to create neutral trials. Including neutral trials is common in the numerical Stroop task and serves to separate facilitation and interference effects by contrasting congruent trials and incongruent trials with neutral ones, respectively (Szűcs & Soltész, 2007). However, trial manipulations in go/nogo and Flanker tasks do not allow to include neutral trials. In order to examine, how the introduced go/nogo manipulation affected facilitation and interference effects in our experiment, we implemented a mass univariate analysis and compared neutral trials with congruent and incongruent ones. The results showed that in the go condition congruent trials were more positive than neutral ones on centro-parietal electrode sites around 325-350 ms. As opposed, in the nogo condition, incongruent trials were more negative than neutral ones around 360-420 ms starting at fronto-central electrode site and reaching to centro-parietal electrode sites as time progressed. These results show that interference and facilitation effects may be altered by differences in cognitive control processes required by go versus nogo trials. More specifically, the conventional peak analysis on the inhibition-related N2 component showed that on incongruent trials interference from physical size is inhibited. The mass univariate analysis advanced this result by showing that when incongruent nogo trials are compared to neutral nogo trials, a stronger and longer inhibitory process is induced. It is likely that the larger negativity on incongruent trials between 360-420 ms reflects an additional inhibitory process or that it reflects response inhibition which the conventional peak analysis was not able to detect. On go trials an opposite effect was observed. Go trials do not require the activation of response inhibition which seems to create an extenuating condition for the integration of physical and numerical size. Thus, we suggest that the larger positivity on the go congruent trials reflect a facilitatory effect induced by the congruent information provided by the physical size. In short, we found evidence for more pronounced inhibitory processes on nogo trials and evidence for facilitation from physical size on go trials. Including neutral trials and performing a mass univariate analysis was necessary to discover these effects, however, further research is required to systematically pinpoint the exact nature of these processes.

When looking at the categorization-related P3 component on go trials with small and large numerical distances, we found that peak latencies were earlier for large compared to small numerical distances, and for congruent than for neutral and for incongruent trials. These findings are in line with Gebuis et al. (2010) who showed numerical distance and congruency effects on the peak latency of the P3 amplitude. Their study, however, did not include neutral trials so it was not possible to separate interference and facilitatory effects of physical size. Our results show that the interaction of physical and numerical size on categorization-related P3 component goes both ways. The P3 latency appears earlier for congruent than for neutral trials, and it also appears earlier for neutral trials than for incongruent trials. Furthermore, we could not show any effect on the P3 peak amplitude that Gebuis et al. (2010) interpreted as a marker of cognitive load (Kok, 2001). The lack of congruency and distance effects on the P3 peak amplitude is probably due to differences in the cognitive load requirements between the classical numerical Stroop and the current paradigm as in the latter one no choice response was required from the participants.

It is important to note that in the current literature it is still debated what effect trial frequency has on the N2 and P3 components. Even though some studies reported larger ERPs on nogo than on go trials irrespective of trial frequency, the difference in amplitude decreased when, for example, go and nogo trials were equally frequent (Bruin & Wijers, 2002; Lavric et al., 2004). Other studies even found a reversed effect when nogo trials were more frequent than go trials (Enriquez-Geppert et al., 2010; Nieuwenhuis et al., 2003). As the role of the proportion of nogo trials is a matter of ongoing debate and because we wanted to test a novel hybrid paradigm, we decided to alternate frequent go trials with infrequent nogo trials. The classical go/nogo ERP effects are the most likely to be found in this case. As follows, it is unclear whether the observed go/nogo ERP effects in the current study reflect response inhibition required on nogo versus go trials or it reflects response conflict created by the unequal ratio of the go and nogo trials.

Furthermore, as mentioned before, recent electrophysiological studies on inhibition have been focusing on disentangling and reliably identifying neural markers specific to different types of inhibition. Such studies implement hybrid inhibition tasks and aim to link conceptually distinguishable inhibitory control functions, such as interference inhibition and response inhibition, to distinct electrophysiological signatures, such as the N2 and the P3. Xie et al. (2017) designed a hybrid Flanker task to elucidate the neural distinction between three types of inhibition: response inhibition, cognitive inhibition (suppressing an irrelevant rule), interference inhibition (suppressing an irrelevant stimulus). They proposed that the N2 is a marker of suppressing irrelevant stimulus. Groom & Cragg (2015) reached similar conclusions when they showed that the N2 is possibly a marker of response conflict which is created by congruent and incongruent stimulus features. One limitation of their study is, however, that they did not keep the overall frequency of the congruent and incongruent trials equal. Therefore, it was not possible to determine whether differences on the N2 amplitude reflect the unequal frequency or the different stimulus features between congruent and incongruent trials. We circumvented this problem by keeping the overall frequency of congruent, neutral and incongruent trials equal throughout the experiment. Thus, the current findings on the N2 amplitude do not reflect differences in trial frequency across the congruency conditions. Another advantage of the current study is that concurrently manipulating two stimulus features (numerical and physical size) made it possible to create neutral trials and therefore to elucidate whether the N2 amplitude is differentially modulated by congruent and neutral trials. This is not possible in Flanker tasks and classical go/nogo tasks because in these paradigms only one stimulus aspect is manipulated. In the current task, the similarity between congruent and neutral trials is that there is no conflict between physical and numerical size even though on congruent trials both stimulus features are manipulated while on neutral trials only numerical size is manipulated. If the N2 is indeed a marker of conflict created by relevant and irrelevant stimulus features – as put forward by Groom & Cragg (2015) – then more negative N2 is expected only on incongruent trials compared to congruent and neutral trials while these latter two are not expected to differ from each other. The results of the current study nicely mimicked this constellation which provides an additional proof that the N2 reflects interference inhibition and is not modulated by facilitatory effects. Moreover, the same pattern of effect was found on go and nogo trials. The current results have significant implications for the field of inhibition research that investigates the exact mechanisms underlying the N2 component and shows that the current paradigm may be a suitable to investigate open questions about distinct neutral markers of different inhibitory processes as well as more specific one in the field of numerical cognition such as how such subprocesses of inhibition contribute to numerical processing in normally developing children and children with mathematical learning disabilities.

Taken together, the current study investigated the electrophysiological correlates of the interaction between numerical and physical size in a modified go/nogo numerical Stroop task. The findings show that the interference between physical and numerical size is probably resolved by inhibitory processes and that facilitatory effects may be affected by cognitive control processes required by go versus nogo trials

Discussion

This thesis consists of two studies. Both of them added to the growing body of research investigating the interplay between visual sensory cues and numerical information. The first study aimed to replicate and extend the findings of Gebuis & Reynvoet (2012a) and showed that continuous magnitudes have a complex influence on numerical processing in a non-symbolic numerical comparison task. The second study investigated the influence of physical size on symbolic numerical processing by introducing a novel go/nogo numerical Stroop paradigm that was able to assess whether this interaction is influenced by inhibitory processes. In the following sections I will briefly summarize the findings and discuss perspectives and future directions.

Study 1 – The interplay between continuous visual properties and non-symbolic number revisited: The effect of mixing trials of different types

In this experiment participants viewed pairs of dot arrays and their task was to decide which array was more numerous. Unbeknownst to them, the continuous visual features, such as convex hull, dot diameter, aggregate dot surface and dot density, were manipulated either congruently or incongruently with numerosity. This resulted in four different visual cue manipulation methods, with congruent and incongruent trials within each method. Trials of the different sensory cue manipulation methods were then presented either block-wise or in a mixed fashion to the participants. Also, the visual cues were not informative about numerosity either during the whole task or within one visual cue manipulation method and the same numerosities were presented during both presentation modes. These are important details because the pattern of congruency effects indicated that visual cues had an immense influence on performance. Hence, this effect cannot be accounted for by either the numerical information or correlations between numerosity and the continuous magnitudes. The study had two main findings: (1) it showed that convex hull and dot diameter have an opposite effect on numerical comparisons. In the *convex hull in/congruent* condition participants performed better when the array with more dots had a larger convex hull. However, in the *diameter in/congruent* condition performance was better on incongruent trials, i.e. when the more numerous dot array had smaller sized dots. And (2), while the pattern of congruency effects remained the same, the reliance on visual cues significantly increased when trials of different types were shown in a mixed fashion.

Finding the reversed diameter effect in both presentation methods confirms that even though convex hull and numerosity are combined in a *greater is more* fashion, diameter and numerosity are combined in a *smaller is more* fashion. Furthermore, it also shows that this reversed effect of diameter is a stable one, not merely the by-product of the block-wise presentation mode. As outlined in the introduction, the reversed diameter effect is a curious result because it indicates that not all visual cues are combined with numerosity in a *larger is more* fashion. This is an important result as single cell studies would suggest the opposite. There are two more findings that can attest to the stability of the reversed diameter effect. First, as mentioned before, researchers of the field have put considerable efforts into developing methods to generate non-symbolic numerical stimuli (De Marco & Cutini, 2020; Dehaene et al., 2005; Gebuis & Reynvoet, 2011; Guillaume et al., 2020; Salti et al., 2016). For example, in one of these methods, in the so called Panamath method, dot stimuli are either *correlated* or *anti-correlated* in terms of numerosity and cumulative surface area. As a results, on a *correlated* stimulus pair the more numerous dot array has larger dots and so, they are comparable to the *diameter congruent* trials in the current study. As opposed, on an *anti-correlated* stimulus pair the more numerous dot array has smaller sized dots which renders these trials similar to the *diameter incongruent* ones. Even though studies that used this dot generation protocol did not focus on the effect of dot diameter on numerical comparisons, their results are in line with the notion that dot diameter and numerosity are combined in a *smaller is more* fashion (Clayton et al., 2015; Norris et al., 2019). Second, in another study Gebuis & Reynvoet (2012b) investigated the influence of continuous magnitudes on numerical judgments in a number line estimation task. They showed participants single dot arrays whose task was to estimate the number of dots in the array and indicate their estimate on the number line. Gebuis & Reynvoet (2012b) were interested in how correlations between numerosity and visual cues affect numerical estimations. To this end, they divided the stimuli of each numerosity into two categories: e.g. stimuli that had a larger than average dot size, and stimuli that had a smaller than average dot size. Their results are also in accord with the reversed diameter effect as it showed that participants gave larger estimates when the array had a smaller than average dot size. There is only one study that found a positive relationship between item diameter and numerosity. Salti et al. (2016) investigated the effect of continuous magnitudes on numerical comparisons in the subitizing range. Subitizing is the ability to quickly and effortlessly enumerate 2-4 objects without the need to count them. Their results indicated that in the subitizing range an increase in dot diameter increased rather than decreased perceived numerosity. Thus, it seems that within the subitizing range there is a positive relationship between dot diameter and numerosity and this relationship is reversed in the estimation range. In sum, it is possible that dot diameter and numerosity are combined in a greater is more fashion in the subitizing range as single cell studies would suggest. This effect is then reversed in the estimation range (>4) which may be a result of learning through past experience. The possible role of learning about the relationship between visual cues and numerosity has been raised by other researchers as well (Gebuis et al., 2016; Leibovich, Katzin, et al., 2016; Mix et al., 2002). Indeed, in natural scenes there tends to be a negative correlation between item number and item size: if there are two baskets of the same size and one of them is filled with apples while the other one is filled with apricots, then the basket with apricots must contain more pieces of fruits.

The second main finding of Study 1 showed that the reliance on continuous magnitudes increased when trials of different types were shown in a mixed fashion. This finding is also in accord with other studies showing that influence of visual cues is not static but may be altered by changes in instruction, task difficulty, stimulus duration and task context (Leibovich et al., 2015; Leibovich, Kallai, et al., 2016; Leibovich, Katzin, et al., 2016; Leibovich-Raveh et al., 2018). The present work discusses two different accounts that may provide an explanation for the interplay between continuous magnitudes and numerical information.

The inhibition account would suggest that during mixed presentation mode, when trials of different types are presented together, it is more difficult to inhibit the visual characteristics than when trials of different types are shown in a block-wise fashion (Leibovich, Kallai, et al., 2016; Leibovich, Katzin, et al., 2016). During blocked presentation mode participants were presented with either congruent or incongruent trials of the same visual cue manipulation method. For example, in the convex hull in/congruent condition participants were presented with pairs of dot arrays that differ in convex hull to a large extent. As such, after a few trials they may become aware of the decorrelation between this visual cue and numerosity and inhibit the reliance on it. As opposed, during mixed presentation mode, trials of different types are shown in a random order. Again, after a few trials they may become aware about the decorrelation between numerosity and visual cues, but in that scenario, it is more difficult to inhibit all visual cues at once, than inhibiting one visual cue during the block-wise presentation mode which leads to the observed effect. It is also possible that due to the random order of different trials types they are not aware of the decorrelation between numerosity and visual cues and this is why they inhibit their effects less than during the block-wise presentation mode.

The idea of sensory integration suggests that the visual cues are weighed and integrated together in order to arrive at a numerosity estimate when comparing dot arrays (Gebuis et al., 2016; Gevers et al., 2016). Thus, this account would suggest that the larger congruency effects during mixed presentation are observed because here larger weights are given to the visual cues then during block-wise presentation. Following the logic described above, participants may realize the decorrelation sooner during block-wise presentation mode which leads to the smaller weights during this condition. In the framework of the integration account, the difference in congruency effects between blocked and mixed presentation mode shows that the weights given to the visual cues are not static, but rather adaptive and depend on experience based on previous trials, even very recent ones. As mentioned before, this finding is also in accord with other studies showing that changes in the influence of continuous visual cues on numerical processing were induced by changes in instruction, task difficulty, stimulus duration and task context (Leibovich et al., 2015; Leibovich, Kallai, et al., 2016; Leibovich, Katzin, et al., 2016; Leibovich-Raveh et al., 2018). These results were interpreted as evidence that the weights given to the continuous magnitudes depends on various top-down and bottom-up processes and as well as on their interaction. However, one might argue that the changes they made to the tasks did not affect the weights per se, but rather they induced changes in processing on a more global level, e.g. in attentional resources or cognitive load. In Study 1, however, the increase in the influence of continuous magnitudes was induced by a very subtle change because the only difference between the two presentation modes was merely changing the trial history. This result, along with the one by Odic et al. (2014), is strong evidence that – if continuous magnitudes are indeed weighed and integrated together – then the weights are not static, but flexible and adaptive or more specifically, they may be adjusted by previous experience. A similar idea has been raised by Petzschner et al. (2015) who proposed a unifying Bayesian framework on magnitude estimation that formally incorporates the influence of prior experience.

It is also important to point out, that the pattern of congruency effects showed an opposite effect of convex hull and dot diameter on numerical comparisons. For the human visual system, it is more sensible to utilize not only convex hull and diameter per se, but also their relationship to each other. The nature of their relationship, e.g. positive or negative, might serve as an evolutionary distinction between depth perception and the processing of large approximate numerosities, when the fast enumeration (subitizing) is not possible. In natural scenes, when the distance between the observer and a set of items changes, it results in a positive relationship between convex hull and diameter. For example, when we see cookies placed on a plate in the cafeteria, it is first translated into a retinal image with a certain convex hull (largest contour around the cookies) and diameter (cookie size). After selecting the plate with more cookies, we will approach it and as we get closer the retinal image of the cookies will grow – including their convex hull and diameter. Thus, decreasing the distance between the observer and the cookies results in larger convex hull and larger diameter while the number of cookies remains the same. In sum, when convex hull and diameter have a positive relationship (they concurrently increase or decrease), it may serve as a cue for change in distance and no change in numerosity. However, to increase the number of cookies on the plate, we either have to put more cookies on the plate (convex hull increases), or we need to replace them with smaller ones (diameter decreases). This example shows that in order to process numerosity it is also essential to process convex hull and diameter together and not in separation.

In sum, Study 1 showed that continuous magnitudes have robust and stable effect on numerical processing, however the extent of their influence may be altered by previous experimental trials. The nature if this effect remains unclear, it may be supported by either sensory integration or inhibition.

Study 2 – Electrophysiological correlates of the interaction of physical and numerical size in symbolic number processing: New insights from a hybrid go/nogo numerical Stroop task

The second study turned to symbolic numbers to investigate the interplay between continuous sensory cues and numerical information. More specifically, the goal of this study was to assess the role of inhibition in the interaction of physical size and numerical size in the numerical Stroop paradigm while examining related electrophysiological components. To this end, a novel hybrid task, the go/nogo numerical Stroop task was introduced to the participants while electrophysiological data were acquired from them. During this task, they were presented with Arabic number pairs with varying numerical distances. To create a Stroop effect, the physical size of the numbers was manipulated which resulted in congruent, neutral and incongruent trials. In order to introduce the go/nogo manipulation, participants were asked to press a button when the number on the one side was numerically lager and refrain from responding when the number on the other side was larger. The study had two main findings that directly relate to the question about the interplay between continuous sensory cues and numerical information.

One, analysing the congruency effects on the inhibition-related N2 and P3 ERP components showed an effect of physical size manipulation on the N2 but not on the P3 component. Incongruent trials were more negative than congruent and neutral trials while these latter two did not differ from one another. Since the N2 component was measured as the peak amplitude between 250-350 ms, this pattern of results shows that the physical and numerical size of the digits as well as their relation to each other is processed prior to this peak. Furthermore, the enhancement on the N2 component was only visible on the incongruent trials when the irrelevant stimulus dimension (physical size) is misleading and interferes with the relevant stimulus dimension (numerical size). In the case of neutral and congruent trials, however, when physical size does not interfere with the numerical size, no enhancement was observed. Thus, it seems that when physical size interferes with numerical size – as in the case of incongruent trials – this interference triggers inhibitory processes that occur between 250-350 ms. Additionally, there was no evidence for the interaction of physical and numerical size on the P3 component. This is implicated by the lack of congruency effect on this component which indicates that it was not modulated by physical size manipulations. Taken together, such a pattern of results implicates that misleading information from physical size induces inhibition at a very early stage, as early as 250-350 ms.

Two, including neutral trials into the go/nogo numerical Stroop task made it possible to separate interference and facilitation effects. These effects are commonly observed in the original version of the numerical Stroop task but they are not investigated in other hybrid go/nogo paradigms because trial manipulations in those experiment do not allow creating neutral trials. Thus, in order to investigate how the go/nogo manipulation affected the well-known interference and facilitation effects, congruent and incongruent trials were contrasted with neutral ones. Indeed, the mass univariate analysis showed that on centro-parietal electrode sites congruent trials in the go condition were more positive than neutral trials between 325-350 ms. As opposed, incongruent trials in the nogo condition were more negative than neutral trials between 360-420 ms starting at fronto-central electrode site reaching to centro-parietal electrode site as time progressed. This pattern of results shows that not only interference but also facilitation effect can be observed in the hybrid go/nogo numerical Stroop task. These results also show that interference and facilitation effects are altered by the different cognitive control processes that are required by go versus nogo trials. The N2 enhancement on incongruent trials revealed by the conventional peak analysis indicates that incongruent information from physical size induces interference in both go and nogo trials that is resolved by inhibitory processes. The mass univariate analysis advanced this finding by showing an additional negativity on incongruent trials compared to neutral trials in the nogo condition 360-420 ms. This effect may reflect an additional inhibitory process that the peak analysis was not able to detect. Indeed, on these trials not only interference from physical size but also motor response had to be inhibited. As the time course of this effect reflects the transition between the interference-inhibition related N2 and the response-inhibition related P3 component, it is difficult to speculate at this point whether the effect reflects a prolonged inhibition of interference from physical size on the incongruent trials or rather a difference in response-inhibition between these trials and the neutral trials.

These are important findings for two reasons. One, previous studies that investigated the role of inhibition in numerical processing were not designed to establish a direct link between these two processes. This study is the first to provide robust evidence for the involvement of inhibitory processes in numerical processing. Investigating inhibition in the context of numerical cognition is of high importance because poor inhibition skills have been implicated in developmental dyscalculia (Szűcs, Devine, et al., 2013; Wang et al., 2012). Therefore, it is possible that there may be differential patterns between normally developing children and adults as well as between normally developing children and those with developmental dyscalculia. The paradigm introduced in this study presents as an effective tool to investigate these possibilities.

Two, as mentioned before, the facilitation and interference effects revealed by the mass univariate analysis implicate that the contribution of these processes in numerical comparisons can be altered by differences of cognitive control processes that are required by the go and nogo trials. Whether symbolic numbers and non-symbolic numerosities are represented by the same systems is still a question of ongoing debate (Krajcsi et al., 2016; Leibovich & Ansari, 2016; Marinova et al., 2020; Núñez, 2017; Reynvoet & Sasanguie, 2016). However, studies with non-symbolic numerosities have shown that the involvement of inhibition and integration may be altered in a flexible and adaptive way (Leibovich et al., 2015; Leibovich, Kallai, et al., 2016; Leibovich, Katzin, et al., 2016; Leibovich-Raveh et al., 2018). Thus, it seems that even if symbolic and non-symbolic numbers are represented by different systems, these results imply that the interplay between numerical and physical information are both mediated in a very flexible and adaptive manner.

As referred to it in the introduction, research on numerical cognition is not autotelic but due to its role in understanding dyscalculia it also has implications for education and legislative processes in general. For example, one important question is what contributes to increasing accuracy in numerical cognition with age and education in children. An overarching study by Piazza et al. (2018) contrasted the classical view, the sharpening hypothesis, with an alternative possibility, the filtering hypotheses. They included numerate and non-numerate children as well as adults who either received or did not receive formal education in order to differentiate between education and maturation effects. According to the sharpening hypothesis, the mental representations on number become increasingly accurate which contributes to the increase in accuracy in numerical discrimination tasks. As opposed, according to the filtering hypothesis, children become better at filtering out non-numerical information. The two theoretical frameworks have distinct predictions regarding congruency effects in non-symbolic numbers. The sharpening hypothesis would suggest a reduction in error rates with age and education, the filtering hypothesis rather suggests a decrease in congruency effect by an increase in accuracy on incongruent trials. Theoretically, it is also possible that the filtering may become so successful that it leads to a decrease in performance on congruent trials where non-numerical parameters would correlate, and as such, potentially facilitate numerical processing. The findings by Piazza et al. (2018) are in accord with the filtering hypothesis. They showed that the increased ability to filter out irrelevant information and to focus on number is an important factor in numerical development. It is important to note, however, that they left open the question about the nature of the filtering process but named inhibition as one available option.

The current study on symbolic numbers adds to these findings by showing that inhibition of physical size is key in symbolic numerical comparisons. Furthermore, it seems that physical size is inhibited when it is not only irrelevant but also interferes with numerical information (incongruent trials). At the same time no evidence for the direct involvement of inhibition was found when physical size was irrelevant but did not interfere with physical size (congruent trials). This pattern of results is difficult to reconcile with the filtering hypothesis. It seems that, at least in non-dyscalculic young adults with formal education, numerical and physical information as well as their relation to one another is processed very early and physical information is only filtered out (inhibited) when it is misleading.

General conclusion

The current work consists of two studies both of which investigated the processes underlying the interplay between numerical and sensory information. Study 1 focused on non-symbolic numbers and showed that sensory cues have a robust and stable effect on numerical comparisons. Study 2, however, turned to symbolic number processing and provided direct evidence for the involvement of inhibition in the interaction of continuous sensory cues and numerical information. Moreover, the novel hybrid paradigm introduced in Study 2 may suitable to investigate the contribution of these processes across different developmental stages and numeracy levels, such as individuals with and without dyscalculia.

References

Agrillo, C., Dadda, M., Serena, G., & Bisazza, A. (2008). Do fish count? Spontaneous discrimination of quantity in female mosquitofish. *Animal Cognition*, *11*(3), 495–503.

Bruin, K. J., & Wijers, A. A. (2002). Inhibition, response mode, and stimulus probability: A comparative event-related potential study. *Clinical Neurophysiology*, *113*(7), 1172–1182. https://doi.org/10.1016/S1388-2457(02)00141-4

Bruin, K. J., Wijers, A. A., & Van Staveren, A. S. J. (2001). Response priming in a go/nogo task: Do we have to explain the go/nogo N2 effect in terms of response activation instead of inhibition? *Clinical Neurophysiology*, *112*(9), 1660–1671. https://doi.org/10.1016/S1388-2457(01)00601-0

Bryce, D., Szücs, D., Soltész, F., & Whitebread, D. (2011). The development of inhibitory control: An averaged and single-trial Lateralized Readiness Potential study. *Neuroimage*, *57*(3), 671–685. https://doi.org/10.1016/j.neuroimage.2010.12.006

Brydges, C. R., Anderson, M., Reid, C. L., & Fox, A. M. (2013). Maturation of cognitive control: Delineating response inhibition and interference suppression. *PloS One*, *8*(7), e69826.

Brydges, C. R., Clunies-Ross, K., Clohessy, M., Lo, Z. L., Nguyen, A., Rousset, C., Whitelaw, P., Yeap, Y. J., & Fox, A. M. (2012). Dissociable components of cognitive control: An event-related potential (ERP) study of response inhibition and interference suppression. *PloS One*, *7*(3), e34482.

Butterworth, B. (2003). *Dyscalculia screener*. Citeseer.

Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. *Trends in Cognitive Sciences*, *13*(2), 83–91. https://doi.org/10.1016/j.tics.2008.11.007

Cappelletti, M., Chamberlain, R., Freeman, E. D., Kanai, R., Butterworth, B., Price, C. J., & Rees, G. (2014). Commonalities for numerical and continuous quantity skills at temporo-parietal junction. *Journal of Cognitive Neuroscience*, *26*(5), 986–999.

Cappelletti, M., Freeman, E. D., & Butterworth, B. L. (2011). Time processing in dyscalculia. *Frontiers in Psychology*, *2*(364), 2.

Cappelletti, M., Pikkat, H., Upstill, E., Speekenbrink, M., & Walsh, V. (2015). Learning to integrate versus inhibiting information is modulated by age. *Journal of Neuroscience*, *35*(5), 2213–2225. https://doi.org/10.1523/JNEUROSCI.1018-14.2015

Cappelletti, M., & Price, C. J. (2014). Residual number processing in dyscalculia. *NeuroImage: Clinical*, *4*, 18–28.

Clayton, S., & Gilmore, C. (2015). Inhibition in dot comparison tasks. *Zdm*, *47*(5), 759–770. https://doi.org/10.1007/s11858-014-0655-2

Clayton, S., Gilmore, C., & Inglis, M. (2015). Dot comparison stimuli are not all alike: The effect of different visual controls on ANS measurement. *Acta Psychologica*, *161*, 177–184.

Clayton, S., Inglis, M., & Gilmore, C. (2019). Developmental differences in approaches to nonsymbolic comparison tasks. *Quarterly Journal of Experimental Psychology*, *72*(3), 436–445.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences. Abingdon*. United Kingdom: Routledge.

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, *3*(2), 63–68. https://doi.org/10.1016/j.tine.2013.12.001

De Marco, D., & Cutini, S. (2020). Introducing CUSTOM: A customized, ultraprecise, standardization-oriented, multipurpose algorithm for generating nonsymbolic number stimuli. *Behavior Research Methods*, 1–10. https://doi.org/10.3758/s13428-019-01332-z

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1), 1–42. https://doi.org/10.1016/0010-0277(92)90049-N

Dehaene, S. (1996). The Organization of Brain Activations in Number Comparison: Event-Related Potentials and the Additive-Factors Method. *Journal of Cognitive Neuroscience*, *8*(1), 47–68. https://doi.org/10.1162/jocn.1996.8.1.47

Dehaene, S. (2001). Précis of the number sense. *Mind & Language*, *16*(1), 16–36.

Dehaene, S. (2003). The neural basis of the Weber–Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences, 7*(4), 145–147. https://doi.org/10.1016/S1364-6613(03)00055-X

Dehaene, S. (2007). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. *Sensorimotor Foundations of Higher Cognition*, *22*, 527–574.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371. https://doi.org/10.1037/0096-3445.122.3.371

Dehaene, S., & Changeux, J.-P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience*, *5*(4), 390–407. https://doi.org/10.1162/jocn.1993.5.4.390

Dehaene, S., Izard, V., & Piazza, M. (2005). *Control over non-numerical parameters in numerosity experiments*.

Dehaene, S., Izard, V., Pica, P., & Spelke, E. (2006). Core knowledge of geometry in an Amazonian indigene group. *Science*, *311*(5759), 381–384.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science*, *320*(5880), 1217–1220.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

DeWind, N. K., Adams, G. K., Platt, M. L., & Brannon, E. M. (2015). Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition*, *142*, 247–265. https://doi.org/10.1016/j.cognition.2015.05.016

Donchin, E. (1981). Presidential address, 1980. *Surprise!... Surprise? Psychophysiology*, *18*, 493–513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x

Donkers, F. C., & Van Boxtel, G. J. (2004). The N2 in go/no-go tasks reflects conflict monitoring not response inhibition. *Brain and Cognition*, *56*(2), 165–176. https://doi.org/10.1016/j.bandc.2004.04.005

Dormal, V., & Pesenti, M. (2009). Common and specific contributions of the intraparietal sulci to numerosity and length processing. *Human Brain Mapping*, *30*(8), 2466–2476.

Eiselt, A.-K., & Nieder, A. (2013). Representation of abstract quantitative rules applied to spatial and numerical magnitudes in primate prefrontal cortex. *Journal of Neuroscience*, *33*(17), 7526–7534.

Enriquez-Geppert, S., Konrad, C., Pantev, C., & Huster, R. J. (2010). Conflict and inhibition differentially affect the N200/P300 complex in a combined go/nogo and stop-signal task. *Neuroimage*, *51*(2), 877–887. https://doi.org/10.1016/j.neuroimage.2010.02.043

Falkenstein, M., Hoormann, J., & Hohnsbein, J. (1999). ERP components in Go/Nogo tasks and their relation to inhibition. *Acta Psychologica*, *101*(2–3), 267–291. https://doi.org/10.1016/S0001-6918(99)00008-6

Fallgatter, A. J., Bartsch, A. J., & Herrmann, M. J. (2002). Electrophysiological measurements of anterior cingulate function. *Journal of Neural Transmission*, *109*(5–6), 977–988. https://doi.org/10.1007/s007020200080

Fallgatter, A. J., Bartsch, A. J., Strik, W. K., Mueller, T. J., Eisenack, S. S., Neuhauser, B., Aranda, D., & Herrmann, M. J. (2001). Test-retest reliability of electrophysiological parameters related to cognitive motor control. *Clinical Neurophysiology*, *112*(1), 198–204. https://doi.org/10.1016/S1388-2457(00)00505-8

Fallgatter, A. J., Brandeis, D., & Strik, W. K. (1997). A robust assessment of the NoGo-anteriorisation of P300 microstates in a cued Continuous Performance Test. *Brain Topography*, *9*(4), 295–302. https://doi.org/10.1007/BF01464484

Fallgatter, A. J., Eisenack, S. S., Neuhauser, B., Aranda, D., Scheuerpflug, P., & Herrmann, M. J. (2000). Stability of late event-related potentials: Topographical descriptors of motor control compared with the P300 amplitude. *Brain Topography*, *12*(4), 255–261. https://doi.org/10.1023/A:1023403420864

Fallgatter, A. J., & Strik, W. K. (1999). The NoGo-anteriorization as a neurophysiological standard-index for cognitive response control. *International Journal of Psychophysiology*, *32*(3), 233–238. https://doi.org/10.1016/S0167-8760(99)00018-5

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology*, *123*, 53–72.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. https://doi.org/10.1016/j.tics.2004.05.002

Feigenson, L., Libertus, M. E., & Halberda, J. (2013). Links between the intuitive sense of number and formal mathematics ability. *Child Development Perspectives*, *7*(2), 74–79.

Fuhs, M. W., & McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: Contributions of inhibitory control. *Developmental Science*, *16*(1), 136–148. https://doi.org/10.1111/desc.12013

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, *4*(2), 59–65. https://doi.org/10.1016/S1364-6613(99)01424-2

Gebuis, T., Kadosh, R. C., & Gevers, W. (2016). Sensory-integration system rather than approximate number system underlies numerosity processing: A critical review. *Acta Psychologica*, *171*, 17–35. https://doi.org/10.1016/j.actpsy.2016.09.003

Gebuis, T., Kenemans, J. L., de Haan, E. H., & van der Smagt, M. J. (2010). Conflict processing of symbolic and non-symbolic numerosity. *Neuropsychologia*, *48*(2), 394–401. https://doi.org/10.1016/j.neuropsychologia.2009.09.027

Gebuis, T., & Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behavior Research Methods*, *43*(4), 981–986. https://doi.org/10.3758/s13428-011-0097-5

Gebuis, T., & Reynvoet, B. (2012a). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General*, *141*(4), 642–648. https://doi.org/10.1037/a0026218

Gebuis, T., & Reynvoet, B. (2012b). The role of visual information in numerosity estimation. *PloS One*, *7*(5), e37426.

Gebuis, T., & Reynvoet, B. (2013). The neural mechanisms underlying passive and active processing of numerosity. *NeuroImage*, *70*, 301–307. https://doi.org/10.1016/j.neuroimage.2012.12.048

Gevers, W., Kadosh, R. C., & Gebuis, T. (2016). Sensory integration theory: An alternative to the approximate number system. In *Continuous issues in numerical cognition* (pp. 405–418). Elsevier.

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., Simms, V., & Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PloS One*, *8*(6), e67374. https://doi.org/10.1371/journal.pone.0067374

Gilmore, C., Cragg, L., Hogan, G., & Inglis, M. (2016). Congruency effects in dot comparison tasks: Convex hull is more important than dot area. *Journal of Cognitive Psychology*, *28*(8), 923–931.

Gomez, A., Piazza, M., Jobert, A., Dehaene-Lambertz, G., Dehaene, S., & Huron, C. (2015). Mathematical difficulties in developmental coordination disorder: Symbolic and nonsymbolic number processing. *Research in Developmental Disabilities*, *43*, 167–178.

Groom, M. J., & Cragg, L. (2015). Differential modulation of the N2 and P3 event-related potentials by response conflict and inhibition. *Brain and Cognition*, *97*, 1–9. https://doi.org/10.1016/j.bandc.2015.04.004

Guillaume, M., Schiltz, C., & Van Rinsveld, A. (2020). NASCO: A new method and program to generate dot arrays for non-symbolic number comparison tasks. *Journal of Numerical Cognition*, *6*(1), 129–147. https://doi.org/10.5964/jnc.v6i1.231

Heine, A., Wißmann, J., Tamm, S., De Smedt, B., Schneider, M., Stern, E., Verschaffel, L., & Jacobs, A. M. (2013). An electrophysiological investigation of non-symbolic magnitude processing: Numerical distance effects in children with and without mathematical learning disabilities. *Cortex*, *49*(8), 2162–2177. https://doi.org/10.1016/j.cortex.2012.11.009

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

Houdé, O., Pineau, A., Leroux, G., Poirel, N., Perchey, G., Lanoë, C., Lubin, A., Turbelin, M.-R., Rossi, S., & Simon, G. (2011). Functional magnetic resonance imaging study of Piaget's conservation-of-number task in preschool and school-age children: A neo-Piagetian approach. *Journal of Experimental Child Psychology*, *110*(3), 332–346. https://doi.org/10.1016/j.jecp.2011.04.008

Huang, H.-W., Nascimben, M., Wang, Y.-Y., Fong, D.-Y., Tzeng, O. J.-L., & Huang, C.-M. (2021). Which digit is larger? Brain responses to number and size interactions in a numerical Stroop task. *Psychophysiology*, *58*(3), e13744.

Huster, R. J., Eichele, T., Enriquez-Geppert, S., Wollbrink, A., Kugel, H., Konrad, C., & Pantev, C. (2011). Multimodal imaging of functional networks and event-related potentials in performance monitoring. *Neuroimage*, *56*(3), 1588–1597. https://doi.org/10.1016/j.neuroimage.2011.03.039

Hyde, D. C., & Spelke, E. S. (2009). All numbers are not equal: An electrophysiological investigation of small and large number representations. *Journal of Cognitive Neuroscience*, *21*(6), 1039–1053.

Hyde, D. C., & Spelke, E. S. (2011). Neural signatures of number processing in human infants: Evidence for two core systems underlying numerical cognition. *Developmental Science*, *14*(2), 360–371.

Hyde, D. C., & Spelke, E. S. (2012). Spatiotemporal dynamics of processing nonsymbolic number: An event-related potential source localization study. *Human Brain Mapping*, *33*(9), 2189–2203. https://doi.org/10.1002/hbm.21352

Inglis, M., & Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychologica*, *145*, 147–155.

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*(3), 1221–1247. https://doi.org/10.1016/j.cognition.2007.06.004

Jodo, E., & Kayama, Y. (1992). Relation of a negative ERP component to response inhibition in a Go/No-go task. *Electroencephalography and Clinical Neurophysiology*, *82*(6), 477–482. https://doi.org/10.1016/0013-4694(92)90054-L

Jongen, E. M., & Jonkman, L. M. (2008). The developmental pattern of stimulus and response interference in a color-object Stroop task: An ERP study. *BMC Neuroscience*, *9*(1), 1–24. https://doi.org/10.1186/1471-2202-9-82

Kadosh, R. C., 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.

Kadosh, R. C., Kadosh, K. C., 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.

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*(3), 888–898. https://doi.org/10.1016/j.neuroimage.2004.12.041

Klein, E., Suchan, J., Moeller, K., Karnath, H.-O., Knops, A., Wood, G., Nuerk, H.-C., & Willmes, K. (2016). Considering structural connectivity in the triple code model of numerical cognition: Differential connectivity for magnitude processing and arithmetic facts. *Brain Structure and Function*, *221*(2), 979–995. https://doi.org/10.1007/s00429-014-0951-1

Knops, A. (2017). Probing the neural correlates of number processing. *The Neuroscientist*, *23*(3), 264–274. https://doi.org/10.1177/1073858416650153

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*(3), 557–577.

Krajcsi, A., Lengyel, G., & Kojouharova, P. (2016). The source of the symbolic numerical distance and size effects. *Frontiers in Psychology*, *7*, 1795.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science*, *197*(4305), 792–795. https://doi.org/10.1126/science.887923

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

Neuroscience, *20*(9), 2483–2488. https://doi.org/10.1111/j.1460-9568.2004.03683.x

Leibovich, T., & Ansari, D. (2016). The symbol-grounding problem in numerical cognition: A review of theory, evidence, and outstanding questions. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *70*(1), 12.

Leibovich, T., & Ansari, D. (2017). Accumulation of non-numerical evidence during nonsymbolic number processing in the brain: An fMRI study. *Human Brain Mapping*, *38*(10), 4908–4921. https://doi.org/10.1002/hbm.23703

Leibovich, T., & Henik, A. (2014). Comparing performance in discrete and continuous comparison tasks. *The Quarterly Journal of Experimental Psychology*, *67*(5), 899–917. https://doi.org/10.1080/17470218.2013.837940

Leibovich, T., Henik, A., & Salti, M. (2015). Numerosity processing is context driven even in the subitizing range: An fMRI study. *Neuropsychologia*, *77*, 137–147. https://doi.org/10.1016/j.neuropsychologia.2015.08.016

Leibovich, T., Kallai, A. Y., & Itamar, S. (2016). What do we measure when we measure magnitudes? In *Continuous issues in numerical cognition* (pp. 355–373). Elsevier.

Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2016). From ‘sense of number’ to ‘sense of magnitude’ – The role of continuous magnitudes in numerical cognition. *Behavioral and Brain Sciences*, 1–62. https://doi.org/10.1017/S0140525X16000960

Leibovich, T., Vogel, S. E., Henik, A., & Ansari, D. (2016). Asymmetric processing of numerical and nonnumerical magnitudes in the brain: An fMRI study. *Journal of Cognitive Neuroscience*, *28*(1), 166–176.

Leibovich-Raveh, T., Stein, I., Henik, A., & Salti, M. (2018). Number and Continuous Magnitude Processing Depends on Task Goals and Numerosity Ratio. *Journal of Cognition*, *1*(1). https://doi.org/10.5334/joc.22

Lyons, I. M., Ansari, D., & Beilock, S. L. (2015). Qualitatively different coding of symbolic and nonsymbolic numbers in the human brain. *Human Brain Mapping*, *36*(2), 475–488. https://doi.org/10.1002/hbm.22641

Marinova, M., Sasanguie, D., & Reynvoet, B. (2020). Numerals do not need numerosities: Robust evidence for distinct numerical representations for symbolic and non-symbolic numbers. *Psychological Research*, 1–13. https://doi.org/10.1007/s00426-019-01286-z

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science*, *211*(4477), 77–80. https://doi.org/10.1126/science.7444452

Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). Multiple cues for quantification in infancy: Is number one of them? *Psychological Bulletin*, *128*(2), 278.

Moeller, K., Willmes, K., & Klein, E. (2015). A review on functional and structural brain connectivity in numerical cognition. *Frontiers in Human Neuroscience*, *9*, 227. https://doi.org/10.3389/fnhum.2015.00227

Moyer, R. S., & Landauer, T. (1967). Time required for Judgements of Numerical Inequality. *Nature*, *215*(5109), 1519–1520. https://doi.org/10.1038/2151519a0

Nieder, A. (2016). The neuronal code for number. *Nature Reviews Neuroscience*, *17*(6), 366. https://doi.org/10.1038/nrn.2016.40

Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience*, *32*, 185–208. https://doi.org/10.1146/annurev.neuro.051508.135550

Nieder, A., Freedman, D. J., & Miller, E. K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science*, *297*(5587), 1708–1711.

Nieuwenhuis, S., Yeung, N., & Cohen, J. D. (2004). Stimulus modality, perceptual overlap, and the go/no-go N2. *Psychophysiology*, *41*(1), 157–160. https://doi.org/10.1046/j.1469-8986.2003.00128.x

Nieuwenhuis, S., Yeung, N., Van Den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: Effects of response conflict and trial type frequency. *Cognitive, Affective, & Behavioral Neuroscience*, *3*(1), 17–26. https://doi.org/10.3758/CABN.3.1.17

Norris, J. E., Clayton, S., Gilmore, C., Inglis, M., & Castronovo, J. (2019). The measurement of approximate number system acuity across the lifespan is compromised by congruency effects. *Quarterly Journal of Experimental Psychology*, *72*(5), 1037–1046.

Núñez, R. E. (2017). Is there really an evolved capacity for number? *Trends in Cognitive Sciences*, *21*(6), 409–424.

Odic, D., Hock, H., & Halberda, J. (2014). Hysteresis affects approximate number discrimination in young children. *Journal of Experimental Psychology: General*, *143*(1), 255. https://doi.org/10.1037/a0030825

Park, J., DeWind, N. K., Woldorff, M. G., & Brannon, E. M. (2015). Rapid and direct encoding of numerosity in the visual stream. *Cerebral Cortex*, *26*(2), 748–763.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*, 10. https://doi.org/10.3389/neuro.11.010.2008

Peters, L., Polspoel, B., de Beeck, H. O., & De Smedt, B. (2016). Brain activity during arithmetic in symbolic and non-symbolic formats in 9–12 year old children. *Neuropsychologia*, *86*, 19–28. https://doi.org/10.1016/j.neuropsychologia.2016.04.001

Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, *19*(5), 285–293.

Piaget, J. (1954). The child’s conception of number. *Journal of Consulting Psychology*, *18*(1), 76.

Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences*, *14*(12), 542–551.

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, *44*(3), 547–555. https://doi.org/10.1016/j.neuron.2004.10.014

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron*, *53*(2), 293–305. https://doi.org/10.1016/j.neuron.2006.11.022

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*(5695), 499–503.

Pinel, P., Dehaene, S., Riviere, D., & LeBihan, D. (2001). Modulation of parietal activation by semantic distance in a number comparison task. *Neuroimage*, *14*(5), 1013–1026.

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.

Plodowski, A., Swainson, R., Jackson, G. M., Rorden, C., & Jackson, S. R. (2003). Mental representation of number in different numerical forms. *Current Biology*, *13*(23), 2045–2050.

Randall, W. M., & Smith, J. L. (2011). Conflict and inhibition in the cued-Go/NoGo task. *Clinical Neurophysiology*, *122*(12), 2400–2407. https://doi.org/10.1016/j.clinph.2011.05.012

Reynvoet, B., & Sasanguie, D. (2016). The symbol grounding problem revisited: A thorough evaluation of the ANS mapping account and the proposal of an alternative account based on symbol–symbol associations. *Frontiers in Psychology*, *7*, 1581.

Rosenthal, R. (1991). *Meta-Analytic Procedures for Social Research* (Revised, Subsequent). SAGE Publications Inc.

Salti, M., Katzin, N., Katzin, D., Leibovich, T., & Henik, A. (2016). One tamed at a time: A new approach for controlling continuous magnitudes in numerical comparison tasks. *Behavior Research Methods*, 1–8. https://doi.org/10.3758/s13428-016-0772-7

Schwarz, W., & Heinze, H. (1998). On the interaction of numerical and size information in digit comparison: A behavioral and event-related potential study. *Neuropsychologia*, *36*(11), 1167–1179.

Shalev, R. S. (2007). *Prevalence of developmental dyscalculia.*

Shalev, R. S., Auerbach, J., Manor, O., & Gross-Tsur, V. (2000). Developmental dyscalculia: Prevalence and prognosis. *European Child & Adolescent Psychiatry*, *9*(2), S58–S64.

Skagenholt, M., Träff, U., Västfjäll, D., & Skagerlund, K. (2018). Examining the Triple Code Model in numerical cognition: An fMRI study. *PloS One*. https://doi.org/10.1371/journal.pone.0199247

Smets, K., Moors, P., & Reynvoet, B. (2016). Effects of presentation type and visual control in numerosity discrimination: Implications for number processing? *Frontiers in Psychology*, *7*, 66.

Smets, K., Sasanguie, D., Szücs, D., & Reynvoet, B. (2015). The effect of different methods to construct non-symbolic stimuli in numerosity estimation and comparison. *Journal of Cognitive Psychology*, *27*(3), 310–325.

Smith, J. L., Johnstone, S. J., & Barry, R. J. (2006). Effects of pre-stimulus processing on subsequent events in a warned Go/NoGo paradigm: Response preparation, execution and inhibition. *International Journal of Psychophysiology*, *61*(2), 121–133. https://doi.org/10.1016/j.ijpsycho.2005.07.013

Smith, J. L., Johnstone, S. J., & Barry, R. J. (2007). Response priming in the Go/NoGo task: The N2 reflects neither inhibition nor conflict. *Clinical Neurophysiology*, *118*(2), 343–355. https://doi.org/10.1016/j.clinph.2006.09.027

Smith, J. L., Smith, E. A., Provost, A. L., & Heathcote, A. (2010). Sequence effects support the conflict theory of N2 and P3 in the Go/NoGo task. *International Journal of Psychophysiology*, *75*(3), 217–226. https://doi.org/10.1016/j.ijpsycho.2009.11.002

Soltész, F., Goswami, U., White, S., & Szűcs, D. (2011). Executive function effects and numerical development in children: Behavioural and ERP evidence from a numerical Stroop paradigm. *Learning and Individual Differences*, *21*(6), 662–671.

Soltész, F., & Szűcs, D. (2014). Neural adaptation to non-symbolic number and visual shape: An electrophysiological study. *Biological Psychology*, *103*, 203–211. https://doi.org/10.1016/j.biopsycho.2014.09.006

Stoianov, I. (2014). Generative processing underlies the mutual enhancement of arithmetic fluency and math-grounding number sense. *Frontiers in Psychology*, *5*, 1326.

Stoianov, I., & Zorzi, M. (2012). Emergence of a’visual number sense’in hierarchical generative models. *Nature Neuroscience*, *15*(2), 194. https://doi.org/10.1038/nn.2996

Strik, W. K., Fallgatter, A. J., Brandeis, D., & Pascual-Marqui, R. D. (1998). Three-dimensional tomography of event-related potentials during response inhibition: Evidence for phasic frontal lobe activation. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *108*(4), 406–413. https://doi.org/10.1016/S0168-5597(98)00021-5

Szűcs, D., & Csépe, V. (2005). The parietal distance effect appears in both the congenitally blind and matched sighted controls in an acoustic number comparison task. *Neuroscience Letters*, *384*(1–2), 11–16.

Szűcs, D., Devine, A., Soltész, F., Nobes, A., & Gabriel, F. (2013). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex*, *49*(10), 2674–2688.

Szűcs, D., Nobes, A., Devine, A., Gabriel, F. C., & Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. *Frontiers in Psychology*, *4*, 444.

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. (2012). Functional definition of the N450 event-related brain potential marker of conflict processing: A numerical Stroop study. *BMC Neuroscience*, *13*(1), 35. https://doi.org/10.1186/1471-2202-13-35

Szűcs, D., Soltész, F., & White, S. (2009). Motor conflict in Stroop tasks: Direct evidence from single-trial electro-myography and electro-encephalography. *Neuroimage*, *47*(4), 1960–1973.

Tang, J., Critchley, H. D., Glaser, D. E., Dolan, R. J., & Butterworth, B. (2006). Imaging informational conflict: A functional magnetic resonance imaging study of numerical Stroop. *Journal of Cognitive Neuroscience*, *18*(12), 2049–2062. https://doi.org/10.1162/jocn.2006.18.12.2049

Tekok-Kilic, A., Shucard, J. L., & Shucard, D. W. (2001). Stimulus modality and Go/NoGo effects on P3 during parallel visual and auditory continuous performance tasks. *Psychophysiology*, *38*(3), 578–589. https://doi.org/10.1017/S0048577201991279

Tokita, M., & Ishiguchi, A. (2013). Effects of perceptual variables on numerosity comparison in 5–6-year-olds and adults. *Frontiers in Psychology*, *4*, 431.

Tudusciuc, O., & Nieder, A. (2009). Contributions of primate prefrontal and posterior parietal cortices to length and numerosity representation. *Journal of Neurophysiology*, *101*(6), 2984–2994.

Verguts, T., & Fias, W. (2004). Representation of number in animals and humans: A neural model. *Journal of Cognitive Neuroscience*, *16*(9), 1493–1504. https://doi.org/10.1162/0898929042568497

Vuillier, L., Bryce, D., Szűcs, D., & Whitebread, D. (2016). The maturation of interference suppression and response inhibition: ERP analysis of a cued Go/Nogo task. *PloS One*, *11*(11), e0165697.

Wang, L.-C., Tasi, H.-J., & Yang, H.-M. (2012). Cognitive inhibition in students with and without dyslexia and dyscalculia. *Research in Developmental Disabilities*, *33*(5), 1453–1461.

Watanabe, S. (1998). Discrimination of “four” and “two” by pigeons. *The Psychological Record*, *48*(3), 383–391.

Xie, L., Ren, M., Cao, B., & Li, F. (2017). Distinct brain responses to different inhibitions: Evidence from a modified Flanker Task. *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-04907-y

Yao, Y., Du, F., Wang, C., Liu, Y., Weng, J., & Chen, F. (2015). Numerical processing efficiency improved in children using mental abacus: ERP evidence utilizing a numerical Stroop task. *Frontiers in Human Neuroscience*, *9*, 245.

Acknowledgements

Preparing this dissertation has been the greatest challenge in my life so far. It has been a long, colourful and arduous journey filled with both highs and lows and, of course, obstacles along the way. Here I would like to express my appreciation for those who accompanied me on this journey.

First and foremost, I would like to express my sincere gratitude to my supervisor, Annette Kinder, for her continuous guidance and support of my research. It was a great privilege and honour to start something new together that provided me with invaluable experience and insights while her scientific expertise guided me during this process. I would also like to thank her for her warmth, empathy, and great sense of humour which got me through difficult times.

I am also grateful for my colleague, Jacqueline Wissmann for her friendship and companionship over the years. She provided me solace – when I needed that. She cheered me up – when I needed that.

I would like to express my special thanks to Sascha Tamm for his technical support during the collection of electrophysiological data. I would also like to thank my students Uria Eytani and Wiebke Hofmann, as well as our research assistants Kevin Kietzmann-Massierer und Antonia Voss, for their assistance in the data collection.

I would like to extend my gratitude to all my colleagues at the Department of Psychology of Learning for the convivial conversations as well as for the friendly working environment we share together.

I would like to thank my old and new friends in Berlin, Neeta Buhecha, Marina González Gómez, Jess Ruesch, Péter Csákvári, Judit Süveges, József Bruck and Eszter Fojdl for their company during the long and lonely days of lockdown and pandemic. You helped me through these weird, isolated times. Furthermore, I would also like to thank Balázs Knakker for the personal and professional conversations which were just as much insightful and helpful as light-hearted and interesting.

Finally, a heartfelt thanks goes out to my family and friends in Hungary for their constant encouragement to pursue anything that I want in life. It gives me great comfort to know that you are always there for me – through the highs and the lows.

Statutory Declaration

Affidavit in accordance with Section 7 (4) of the joint doctoral regulations for Dr. rer. nat./ Ph.D. the Faculty of Education and Psychology at the Free University of Berlin, published from 8 August 2016:

I hereby declare that this thesis is my own work. I have used only the listed resources and aids. It is on this basis that I wrote the work independently and without (inadmissible) help of third parties. The work contains no material previously published or produced by another party. Furthermore, I declare that the dissertation is neither in part nor in its entirety at another academic university for assessment in a doctoral procedure, nor has it ever been.

Eidesstattliche Erklärung

Eidesstattliche Erklärung nach §7 Abs. 4 der gemeinsamen Promotionsordnung zum Dr. rer. nat./ Ph. D. das Fachbereichs Erziehungswissenschaft und Psychologie der Freien Universität Berlin vom 8. August 2016:

Hiermit erkläre ich, dass diese Arbeit meine eigene Arbeit ist. Ich habe nur die angegebenen Hilfsmittel und Hilfen benutzt. Auf dieser Grundlage habe ich die Arbeit selbstständig und ohne die (unzulässige) Hilfe Dritter verfasst. Die Arbeit enthält kein Material, das zuvor von einer anderen Partei veröffentlicht oder produziert wurde. Weiterhin erkläre ich, dass die Dissertation weder in Teilen noch in ihrer Gesamtheit einer anderen wissenschaftlichen Hochschule zur Begutachtung in einem Promotionsverfahren vorliegt oder vorgelegen hat.

Date/Datum Signature/Unterschrift

Contributors Role

Study 1 has been published and Study 2 is under review as follows:

Study1:

Pekár, J., & Kinder, A. (2020). The interplay between non-symbolic number and its continuous visual properties revisited: Effects of mixing trials of different types. *Quarterly Journal of Experimental Psychology*, *73*(5), 698-710.

Contributors Role – Study 1:

Judit Pekar	Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review and editing, Submission, Correspondence, Revision
Annette Kinder	Resources, Supervision, Writing – review and editing, Revision

Study 2:

Pekár, J., Hofmann, W., Tamm S. & Kinder, A. (under review). Electrophysiological correlates of the interaction of physical and numerical size in symbolic number processing: New insights from a hybrid go/nogo numerical Stroop task

Contributors Role – Study 2:

Judit Pekar	Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review and editing, Submission
Wiebke Hofmann	Data Curation, Project administration, Visualization
Balázs Knakker	Supervision
Sascha Tamm	Software, Supervision
Annette Kinder	Resources, Supervision, Writing – review and editing