+

UNIVERSITÀ DEGLI STUDI DI PADOVA

Sede amministrativa: Università degli Studi di Padova

Dipartimento di Psicologia Generale

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE PSICOLOGICHE

CICLO XXVII

More than space. A new insight into number representation.

Direttore della Scuola: Ch.ma prof.ssa Francesca Peressotti

Supervisore: Ch.ma prof.ssa Mariaelena Tagliabue

Dottorando: Michael Davis Tira

Table Of Contents

Riassunto

Come possiamo rappresentare i numeri e fare calcoli matematici? Questa domanda è l'obiettivo principale del presente lavoro e cade nel campo della cognizione matematica, il quale si interessa dei processi cognitivi e neurologici che sottendono le abilità matematiche.

L'ipotesi della linea numerica mentale (MNL) prevede che i numeri siano rappresentati mentalmente sottoforma di una misura continua (analogica) con valori numerici crescenti da sinistra a destra. La MNL viene considerata uno dei migliori modelli per la rappresentazione mentale dei numeri. Molti studi hanno esaminato la MNL considerando l'effetto SNARC (Spatial-Numerical Association of Response Codes) come prova per una connessione univoca tra spazio e numero. Tuttavia, è stato dimostrato che la rappresentazione mentale di valori piccoli a sinistra e di valori più grandi esiste anche per grandezze diverse dalla numerosità, compresa la durata temporale e la grandezza fisica. Queste osservazioni convergono con l'idea di un sistema dove diverse grandezze (ad esempio tempo, spazio e numerosità) condividono risorse neurali e concettuali, definito sistema generale di elaborazione delle grandezze (GMS). Questo solleva un'importante domanda sulla natura delle informazioni rappresentate lungo la MNL: si tratta esclusivamente di informazioni numeriche?

Il presente lavoro è diviso in 4 capitoli. Il capitolo 1 affronta diversi problemi riguardanti la rappresentazione mentale dei numeri. La ricerca nel campo della cognizione matematica ha una lunga storia e ha fatto notevoli progressi negli ultimi decenni; a volte questo grande volume di dati rende difficile ottenere una visione globale di quello che è lo stato dell'arte. Per questo motivo il Capitolo 1 offrirà una

panoramica dei diversi modelli di rappresentazione mentale dei numeri, sia innati che acquisiti, precisi o approssimati, simbolici o non simbolici. Prima di tutto sono elencate le principali scoperte sulla rappresentazione mentale dei numeri; in secondo luogo verrà presentata una carrellata sulla letteratura che mostra come le rappresentazioni di tempo, spazio, intensità e numero interagiscano tra loro e probabilmente condividano meccanismi di elaborazione; questo fornirà un adeguato contesto teorico necessario alla chiara comprensione dei lavori sperimentali presentati nei capitoli successivi. Una gran quantità di risultati scientifici dimostra che la rappresentazione e l'elaborazione dei numeri siano associate all'attivazione di una rappresentazione di natura spaziale. Una delle posizioni canoniche della cognizione numerica a tal riguardo afferma che la codifica spaziale è una componente imprescindibile della rappresentazione mentale a lungo termine dei numeri. Secondo questa idea, che porta il nome di ipotesi della linea numerica mentale, i numeri sarebbero rappresentati come una linea continua con i numeri più piccoli a sinistra e quelli più grandi a destra. Tuttavia l'origine dell'associazione tra numeri e spazio non è stata ancora totalmente chiarita. Verranno presentati degli studi che dimostrano come la codifica spaziale dei numeri non sia, in effetti, stabile nè necessariamente il risultato di un'associazione a lungo termine, ma al contrario sia una rappresentazione flessibile costruita a partire dalle necessità di elaborazione delle informazioni specifiche per i compiti che ogniuno di noi si trova a svolgere quotidianamente. Inoltre saranno presi in considerazione studi sull'associazione dei numeri con grandezze prive di caratteristiche spaziali.

Nel Capitolo 2 viene presentata una serie di tre studi sperimentali ed in ognuno di essi è stato impiegato un metodo di risposta basato sulla produzione di numerosità. I partecipanti hanno eseguito un compito di aritmetica approssimata su numeri presentati, a seconda dello studio, in notazione simbolica o non simbolica. In tutti gli studi presentati i partecipanti sono stati istruiti ad utilizzare un metodo di risposta caratterizzato dalla produzione di numerosità non simboliche, essi infatti fornivano la risposta al compito specifico nel quale erano impegnati attraverso la produzione, sullo schermo di un computer, di un insieme di punti la cui numerosità era controllata dalla rotazione di una manopola posta davanti ai partecipanti e connessa al computer. Un apposito programma si occupava di registrare il grado di rotazione della manopola ed aggiornare il numero di punti presentati sullo schermo.

Lo studio 1 presenta due esperimenti in cui i partecipanti giudicavano la numerosità media tra due insiemi di punti presentati in sequenza. Nell'Esperimento 1 di questo studio, i partecipanti utilizzavano una scala di numerica di risposta da 0 a20 (scala categorica), mentre nell'Esperimento 2 la risposta è stata data attraverso il metodo di risposta basato sulla produzione di numerosità. I risultati di questo studio hanno mostrato come le risposte siano state fornite secondo un modello di integrazione Average. Questo suggerisce una linearità nella scala risposta per entrambi i metodi usati nel compito di aritmetica approssimativa. Più importante, i due operandi mostravano di esercitare la stessa influenza sulla risposta fornita dai partecipanti, il che esclude un effetto sequenza o recenza legata ai compiti impiegati . Questi due esperimenti sono serviti come strumento di validazione del metodo di risposta basato sulla produzione di numerosità al fine della sua applicazione negli studi successivi.

Lo Studio 2 presenta un esperimento in cui il metodo di risposta basato sulla produzione di numerosità è stato utilizzato per testare l'effetto della forza necessaria a ruotare la manopola usata per portare a termine un compito di aritmetica mentale. In particolare si è verificata l'influenza della variabile Forza sull'effetto denominato

Operational Momentum (OM). L'effetto OM è la tendenza sistematica a sovrastimare i risultati di addizione e a sottovalutare i risultati di sottrazioni in condizioni che impediscono un esatto conteggio. In questo esperimento la forza necessaria per ruotare la manopola è stata manipolata in tre blocchi tra i soggetti. La letteratura ha suggerito che l'effetto OM possa dipendere da una rappresentazione spaziale dei numeri; tuttavia i risultati di questo studio dimostrano che l'eliminazione di un feedback psicomotorio quale la forza richiesta per ruotare la manopola, porta all'annullamento della differenza tra addizioni e sottrazioni. I risultati di questo studio forniscono evidenze sperimentali dell'influenza di una grandezza priva di connotazioni spaziali quale la Forza su un fenomeno di aritmetica mentale come l'effetto OM. Questo risultato è particolarmente interessante considerando che la Forza fosse una variabile interamente irrilevante per lo svolgimento del compito.

Lo Studio 3 presenta un esperimento sul confronto tra quattro diversi effetti classicamente considerati esempi dell'automaticità dell'attivazione di codici spaziali durante l'elaborazione di informazioni numeriche.

Gli effetti che sono stati considerati in questo studio sono l'effetto SNARC, l'effetto distanza, l'effetto di congruenza delle dimensioni e l'effetto OM. L'effetto SNARC: la tendenza ad essere più veloci nel rispondere a numeri piccoli sulla sinistra e a numeri più grandi a destra. L'effetto distanza: il fatto per cui numeri vicini tra loro sono piu difficili da discriminare rispetto a numeri distanti tra loro. L'effetto di congruenza delle dimensioni: il fatto che i numeri sono identificati come maggiori o minori di 5 più rapidamente se la loro dimensione fisica è congruente con la loro grandezza numerica. Ultimo ma non meno importante, l'effetto OM. Tali effetti sono stati testati insieme per indagare i rapporti che li legano con un approccio basato sulle differenze individuali. La presenza di ognuno degli effetti è stata verificata. Al fine di valutare la correlazione tra i vari effetti in esame, è stato calcolato il coefficiente di regressione lineare di ciascun effetto su ognuno dei partecipanti. I risultati di questo studio, anche se non conclusivi, puntano in direzione di una rappresentazione mentale comune tra gli effetti numerici testati (effetto SNARC, effetto di congruenza della dimensione, effetto distanza). L'effetto OM, inoltre, sembra correlare negativamente con l'effetto SNARC, suggerendo una connessione tra i due, ma contraddicendo la teoria della linea numerica mentale.

Nel capitolo 3 si traggono conclusioni sul lavoro sperimentale presentato tenendo conto di diversi quadri esplicativi. Il presente lavoro di ricerca utilizza un metodo di risposta per compiti numerici relativamente poco noto: il metodo di risposta basato sulla produzione di numerosità. Questo metodo presenta una vasta gamma di applicazioni e apre nuovi scenari nel campo della cognizione matematica, fornendo un valido strumento per comprendere nel dettaglio le implicazioni dell'azione nella cognizione matematica. Gli esperimenti qui presentati, inoltre, forniscono indicazioni chiare rispetto al ruolo del feedback psicomotorio con caratteristiche non spaziali in compiti di aritmetica mentale portati a termine attraverso un metodi di risposta basato sulla produzione di numerosità, mettendo così in discussione l'interpretazione classica dell'effetto OM come effetto derivato da una rappresentazione puramente spaziale dei numeri. Considerando che le informazioni riguardanti la forza sono state presentate attraverso un feedback tattile mentre le informazioni numeriche sono state presentate visivamente, tale integrazione tra modalità sensoriali diverse è coerente con l'ipotesi di un sistema generale per le grandezze.

Lo studio 3 confrontando, a nostra conoscenza per la prima volta, diversi effetti legati all'ipotesi della linea numerica mentale, fornisce nuove informazioni sui meccanismi di elaborazioni condivisi a questi classici effetti nel campo della

cognizione matematica. I nostri risultati, anche se non conclusivi, rinnovano la domanda sulla natura della rappresentazione mentale dei numeri.

Parole chiave: Aritmetica mentale; Forza, Operational Momentum; SNARC; Effetto Distanza; Effetto di congruenza delle dimensioni.

Abstract

How do we represent numbers and make mathematical calculations? This question is the main focus of the present work and it falls in the domain of mathematical cognition, the field of knowledge concerned with the cognitive and neurological processes that underline mathematical abilities.

The mental number line, with its analogue left-to-right orientation of growing numerical values, is often regarded as the best candidate to the role of mental representation of numbers. Many studies have examined the so-called mental number line taking the Spatial-Numerical Association of Response Codes (SNARC) effect as evidence for a unique connection between space and number. However, left-to-right orientation has been shown to extend to other dimensions, including duration and physical size. Such observations converge with the notion of a general magnitude system, where different magnitudes share neural and conceptual resources. This rise an important question about the nature of the information represented along the mental number line: is it exclusive to number or not?

The present work is divided in 4 chapters. Chapter 1 addresses several issues in mental representations of numbers. Research in mathematical cognition has a long history and has made considerable progress over the last decades; sometimes this big volume of data makes it difficult to gain a global view of what the state-of-the-art is. For this reason, Chapter 1 will offer an overview of the different mental representations of numbers, whether innate or acquired, precise or approximate, symbolic or non symbolic. On the one hand, the most important insight gained on mental representations of numbers are listed; on the other, literature on the

representation of time, space, number and intensity, related with number representation, are revised to show the similarities between these domains, and how those are indicative of common processing mechanisms. The theoretical background specific for the present work is introduced. A great body of evidences point out that the representation and processing of numbers is associated to an activation of spatial codes. One of the classical view of numerical cognition on this subject states that spatial codes are an imprescindible component of the long-term representation of numerical magnitude information. According to this idea, refereed to as number mental line hypothesis, numbers are systematically associated to spatial codes, as if numerical magnitudes were represented along a spatial continuum with small numbers to the left and large numbers to the right. Nevertheless, the origin of the association between numbers and space is not completely clear to date. Studies will be presented showing that the spatial coding of numbers is not stable and is not necessarily the result of long-term memory associations but, on the contrary, a flexible type of representation built during cognitive processing as the result of task demands and spatial coding preferences. Moreover, studies on the association of numbers with others, non-spatial, magnitudes will be reviewed.

In Chapter 2 a series of three studies are presented, in all of them a numerosity production method of response was used. Participants performed approximate arithmetic task on symbolically presented numbers, they were instructed to respond by the production of a dot pattern, the set size of which was controlled by a rotating knob. Study 1 shows two experiments in which participants judged the average numerosity between two sequentially presented dot patterns. In Experiment 1, the response was given on a 0–20 numerical scale (categorical scaling), and in Experiment 2, the response was given by the production of a dot pattern of the desired numerosity (numerosity production). Data showed that responses were shaped according to an averaging integration model. This suggests the linearity in the response scale of both of the response methods in the approximate arithmetic task. More important, the two operands were found to have the same influence in determining the result. These two experiments served as a validation tool of the numerosity production method of response to be applied in the sequent studies.

Study 2 proposes one experiment using the numerosity production method of response to test the influence of the force required to respond on the Operational Momentum (OM) effect. The OM effect is the finding of a systematic tendency to overestimate the results of addition problems and underestimate the results of subtraction problems under conditions that prevent exact calculation. In this experiment the force required to turn the knob has been manipulated in three between subjects blocks. It has been suggested that the OM effect depend on the spatial representation of numbers; by showing that the elimination of psycho-motor feedback nullifies the difference between addition and subtraction, evidence are provided that the OM effect is modulated by information from a magnitude different from space (required force), even when such information is entirely task-irrelevant.

Study 3 proposes an experiment on the comparison of four different effects classically considered examples of the automatic spatial organization of numerical information. The spatial-numerical association of response codes (SNARC), that is, the tendency to be faster in responding to small numbers on the left and to bigger number on the right. The distance effect, that is, close numbers are more difficult to compare than numbers far apart. The size congruency effect, that is, numbers are identified more rapidly as bigger or smaller than 5 if their physical size is congruent with the correct answer. And last but not least, the OM effect. Those effects have been tested together to investigate the relationship among them with an inter-individual differences approach. The presence of all the effects object of this study was verified in the participants set. Linear regression have been used to calculate the coefficient of each subject for each effect in order to test the correlation between all the effects this study take into consideration. The result of this study, even if not conclusive, point in the direction of a common representational mechanism underling the tested numerical effects (SNARC, size congruency, distance). Moreover, the operational bias seem to have a negative correlation only with the SNARC effect, suggesting a connection between the two, but weekending the mental number line account of those effects

In Chapter 3 conclusions are drawn upon the presented experimental work taking into account different explanatory frameworks. The present research work use a relatively unknown method of response to numerical tasks; the numerosity production method of response. This method shows a wide range of applications and opens new scenarios in mathematical cognition, providing a good instrument to understand in detail the implications of action in mathematical cognition. Moreover, the experiments here presented provide clear indications for a role of non-spatial psycho-motor feedback in arithmetical calculations carried out with the numerosity production method of response, thus challenging the classical interpretation of OM as an effect derived from a purely spatial representation of numbers. Moreover, considering that the force information was presented haptically but numerical information visually, such integration across sensory modalities is consistent with the General magnitude system hypothesis suggesting that representations of magnitudes are multimodal. Study 3 comparing, at our knowledge for the first time, different effects connected to the mental number line hypothesis, provide new insight on the

shared processing undergoing these classical findings of mathematical cognition. Our findings, although not conclusive, renew the question on the nature of the representation of numbers.

Keywords: non-symbolic arithmetic; Force; Operational Momentum; SNARC; Distance effect; size-congruency effect;

Chapter 1

Concepts in mathematical cognition

Introduction

The present work will make use of many concepts taken from the field of mathematical cognition, that is the branch of cognitive science that studies cognitive, developmental and neural bases of numbers and mathematics. To get a first grip on the core question motivating this whole field of research let's ask: what's mathematics? Mathematics is a system to symbolically represent and virtually manipulate quantities; a system so resourceful that when it comes to formally building models of the physical world no conceptual instrument matches mathematics for power and flexibility. Thus, what are the psychological foundations of the human mastery of this system?

I will now briefly review the answers to this question that experimental research on human and non-human animal cognition suggests. Binet (1890) was the first to report about numerosity. He informally investigated the ability of children to compare the numerosity of two presented collections of simple objects. Binet (1890/1969, p. 87) concluded: “if [the child] judges one group more numerous than another, it is because it occupies more space on the paper”. In 1929, De Marchi was the first to use investigate numerical evaluation of collectivities in a proper experiment. According to De Marchi, the evaluation of collectivities refers to “the process by which a perceived aggregate is expressed by numerals in conditions that exclude any possibility of numbering its elements” (De Marchi, 1929/1986, p. 184).

De Marchi acknowledged that variables influencing numerical evaluation—such as the duration of exposure, size of the surface, occupied by the single collectivities, density of the exposed elements (dots), or space and time disposition—could together influence the evaluation in an experiment. Years later, studies on these same collectivities studied by De Marchi, now addressed as numerosity, lead to the conclusion that mathematical abilities are independent from language or other symbol systems. This is known because the ability to estimate quantities and to manipulate arithmetically those estimates exists in non-human animals. A variety of studies have demonstrated that non-human animals, including rats, lions, and various species of primates, have an approximate sense of numbers (for a review, see Dehaene, 1997).together with infants (Feigenson, Dehaene, & Spelke, 2004) and adult humans without any schooling in mathematic (Deaheane, 1997). These findings suggest that numbers and arithmetic thinking is based on a non-verbal system for estimating and manipulating discrete and continuous quantity, a system shared with many non-human species. From this base knowledge, it is reasonable to suppose that the neural substrate for this system was born far back in the evolution in order to capture important properties of the world which individuals must represent to effectively drive their actions.

Terminology

Many of the terms that the reader is going to encounter from here on posses a more specific meaning in the present work than in everyday speech. For this reason, before deepen the discussion a clarification on the terminology that will be in use is in order. Despite number are associated spontaneously, by most western people, with arabic digits, they can also be represented as sequence of words or in an analogue format (I.e dots or any set of objects); from here on I'll address those different number representations as follow: *numerosity*, is used for the mumeric properties of a set of objects; *symbolic codes* or *numerical notations*, identify the system used to present numerical informations (i.e. Arabic numerals, Roman numerals, and number names); and the *internal mental numerical representations*, correspond to entities internal to the subject and concerning both to the system of numerical notation and numerosity. Moreover, according to McCloskey and Macaruso (1995), the term *number* will be used for format-independent aspects of numerical cognition, while the term *numeral* will be reserved to modality-specific representations (i.e. verbal, analogue, arabic numerals). As defined by Dehaene (2009), "*Symbolic arithmetic* deals with how we understand and manipulate numerals and number words" (p. 233). From which it follows that, "*Nonsymbolic arithmetic* is concerned with how we grasp and combine the approximate cardinality or "*numerosity*" of concrete sets of objects (such as visual dots, sounds, and actions)". The term *number system* will be used to refer to the set of entities and causal connections that allow for the arithmetical manipulation of real world quantities representations. Such a system forcibly posses a usefully invertible mapping between those internal representations and the real word entities it represent: the numbers obtained through arithmetic processing correctly refer through the inverse mapping back to the represented reality.

Numerical Estimation and Manipulation in Animals

In considering the literature on numerical estimation and manipulation in animals, the evidence that they estimate and manipulate arithmetically a continuous quantity as time are reviewed. Many animals measure and remember continuous quantities, as has been shown in a variety of experimental paradigms. One of them is the peak procedure. A trial of this paradigm begins with the onset of a stimulus signaling the availability of food at the end of a fixed interval (feed latency). Responses made at or after the interval has elapsed trigger the delivery of food. Response prior to that time did not trigger the delivery of food and have no other consequences. On 20-50% of the trials, food is not delivered even with a response after the feed latency has past (probe trials). The data analyzed in Peak-procedure come from these unrewarded trials. On such trials, the subject begins to respond some while before the interval ends (in anticipation of its ending) and continues for some while after, before abruptly stopping. The interval during which the subject responds circumscribe its subjective estimate of the fixed interval (Church, Meck, & Gibbon, 1994). Interestingly, the trial-to-trial variability in the onsets and offsets of responding is proportional to the latency. That is, the probabilities that the subject begin or stop responding are determined by the proportion of the feed latency that has elapsed. This property of time representation is called scalar variability.

Many animals also count and remember numerosities (Brannon & Roitman, 2003; Church & Meck, 1984; Dehaene, 1997; Dehaene, Dehaene-Lambertz, & Cohen, 1998; Gallistel, 1990; Gallistel & Gelman, 2000).In a common paradigm for assessing counting and numerical memory in animals, the subject must press a lever a target number of time in order to unlock the gate of a feeding box. Pressing too many times gives no penalty. Trying to open the box too soon incurs a 10-second time-out, which the subject must endure before returning to the lever to complete the requisite number of presses (Mechner, 1958; Platt & Johnson, 1971). The number of presses at which subjects are maximally likely to stop pressing and try to enter the alcove, maximizes at or slightly after the required number. Moreover, as the target number gets larger, the variability in the stopping number also gets proportionately greater. That is, also behavior based on number shows scalar variability The fact that behavior based on numerosity exhibits scalar variability just like the scalar variability seen in behavior based on magnitude of continuous quantities like duration suggests that numerosity is represented in the brains of animals by mental magnitudes, that is by analogue dimensions, rather than by discrete symbols like words. Considering a system using discrete entities to represent numerosity there is no reason for the variability of response to be related with the size of the targets number. Thus, the nature of the variability in a target number suggests that the representation of that number is something that behaves like a continuous quantity, a magnitude. A number system is that if the mental entities representing magnitudes in the real word are manipulated in a meaningful way to drove actions. Considerable experimental literature demonstrate that laboratory animals arithmetically manipulate mental magnitudes representing numerosity and duration. Non-human animals have been found able to add, subtract, divide and order durations and numerosities in such a way that their mental operations on subjective quantities enable these animals to behave effectively in relation to the tasks (For reviews, see Sarah T. Boysen & Hallberg, 2000; Elizabeth M. Brannon & Roitman, 2003; S. Dehaene, 1997; Gallistel, 1990; Spelke & Dehaene, 1999).

Comparable Mental Magnitudes for Numerosity and Duration

Gibbon (1977), to explain the generation of mental magnitudes representing durations, had proposed that the ability to perceive duration was formally equivalent to a flow of impulses directed to an accumulator, so that the accumulation grew proportionally to the duration of the flow. When the flow ended, the resulting accumulation, representing the duration of the interval, was read into memory. Meck and Church (1983) pointed out that this mental accumulator model could be modified to make it generate mental magnitudes representing numerosities; they proposed that to get magnitudes representing numerosity, the equivalent of a pulse former (device that, upon receiving a fix amount of a signal outputs a rectangular pulse,) was inserted into the stream of impulses, so that for each count there was a discrete increment in the contents of the accumulator, as happens when a cup of liquid is poured into a graduated container. At the end of the count, the resulting accumulation is read into memory where it represents the numerosity. This discrete version of the accumulation model was originally proposed to explain behavior based on the numerosity of serial events, but it may be generalized to the case where the items to be counted are presented all at once. In the case of a visual array to be enumerated, to each item in the array a unitarian magnitude can be assigned and then accumulated across space, rather than over time. This improved model is the origin of the hypothesis that the mental magnitudes representing duration and the mental magnitudes representing numerosity are essentially the same, differing only in the process mapping the real stimuli to this metal representation. Put another way, both numerosity and duration are represented mentally by continuous magnitudes. Furthermore, Meck and Church (1983) comparing the representation of number and time in the rat, found that the ratio between the standard deviation and the mean (coefficient of variation) of the behavioral measurement in use, was the same for number and time, which is further evidence for the hypothesis that the same system is used in both cases.

Nuerosity as Mental Magnitude in Humans

From a phylogenetic prospective it would be bizarre to think that humans do not share with their more and less distant vertebrate relatives, for example pigeons and chimpanzees, the mental mechanism for representing countable and uncountable quantity by means of a number system. Moyer and Landauer (1967; 1973) first suggested that humans represent integers with mental magnitudes when they discovered what has come to be called the *symbolic distance effect*. When subjects have to judge the numerical order of Arabic numerals as rapidly as possible, their reaction time depend on the relative numerical distance: the greater the distance between the two numbers, the more quickly the task is carried out. Subsequently, Parkman (1971) showed further that the greater the numerical value of the smaller digit, the longer it takes to judge their order, that is the to be called *size effect*. A single law can be used to summarized those two effects: the time to judge the numerical order of two numbers is a function of the ratio of the numerical magnitudes that they represent. It is evident how this can be recollected to a more general low: the discriminability of two magnitudes is a function of their ratio, that is, Weber's law applies to symbolically represented numerical magnitude. The *size* and *distance effects* in human judgments of the ordering of discrete and continuous quantities are robust. They are observed when the numerosities being compared are visual arrays of dots and when they are represented symbolically by Arab numerals (Buckley & Gillman, 1974). Moreover the *symbolic distance* and *size effects* are observed both in the single digit and in the double digit range (Dehaene, Dupoux, & Mehler, 1990; Hinrichs, Yurko, & Hu, 1981). One might think that the facts about which numbers were greater than which were some how stored and simply retrieved. Nevertheless, it take longer to look up the ordering of 2 and 3 than 2 and 6 and this suggests that the comparison that underlies these judgments use mental magnitudes. On this hypothesis, the comparison that mediates the judgment of the numerical ordering of two Arabic numerals uses the same mental magnitudes and the same comparison mechanism used by the non-verbal numerical system that we are assumed to share with many non human animals. Reinforcing this hypothesis, Brannon and Terrace's (2002) using a numerical ordering task on visually presented dot arrays find that the reaction time functions of humans and monkeys are almost exactly the same.

Considering the evidence from the *symbolic size* and *distance effects* that humans represent number with mental magnitudes, it seems likely that they share with the non-human animals a non-verbal counting mechanism that maps from numerosities to the mental magnitudes that represent them. Given that, it should be possible to demonstrate non-verbal counting in humans when verbal counting is suppressed. Presenting subjects with Arabic numerals on a computer screen and asking them to press a key as fast as they could without counting until it felt like they had pressed a number of times equal to the value of the numeral, the results from humans looked very much like the one from pigeons and rats (Whalen, Gallistel, & Gelman, 1999). The mean number of presses increased proportionally to the target number and the standard deviations of the distributions of presses increased in proportion to their mean, so that the coefficient of variation was constant. This finding suggests that subjects could count non-verbally, comparing the mental magnitude thus generated to an other mental magnitude derived from numerals via a learned mapping. Moreover, it implies that, given a numeral, the mental magnitude mapped from that numeral approximate the mental magnitude generated by counting the numerosity signified by that given numeral.

In another task of the same experiment, subjects were asked to observe a dot flashing at irregular intervals. To prevent verbal counting, the rate of flashing (8 per second) was twice as fast as what it's considered the maximum speed for verbal counting (Mandler & Shebo, 1982). As in the first experiment, the mean estimated number increased in proportion to the number of flashes and the standard deviation increased in proportion to the mean estimate. This result show that the mental magnitude generated from a symbol is comparable to the one generated by non-verbal counting, in both cases, the variability in the mapping is scalar. In a control experiments with the same task, Whalen, Gallistel, and Gelman (1999), asked subjects to count aloud their presses (condition *a*) or to say "the" coincident with each press (condition *b*). In all conditions, subjects were asked to press as fast as possible. The variability data from the condition where subjects were required to say "the"confirmed that the coefficient of variation was constant (scalar variability) as in Whalen, Gallistel, & Gelman (1999). Differently, In the condition where subjects counted aloud, one would expect counting errors, as double counts and skips to, main source of variability. Assuming the probability of a counting error as approximately equal at any step of a count, the resulting variability in final counts should be binomial rather than scalar, that is, it should increase in proportion to the square root of the target value, rather than in proportion to the target value. This is what was in fact observed in the aloud counting conditions: the variability was much less than in the non-verbal counting conditions and it was binomial rather than scalar. The different patterns of variability in the counting-aloud and non-verbal counting conditions support the idea that subjects in the non-verbal counting conditions were not subvocally counting. Summarizing, humans are able of non-verbal counting just like non-human are. Moreover, the mental magnitudes generated by non verbal counting appear to be comparable to the ones generated by symbolically presented integers. This may suggests that the linguistic representation of numerosity is based on, and find its meaning in, the non-verbal counting system. Another feature characterizing our representation of numbers beside the *symbolic size* and *distance effects* is the *Spatial-Numerical Association of Response Codes, or SNARC, effect.* That is the finding of faster reaction times for small numbers when responses are made with the left compared to the right hand and the opposite pattern for large numbers. (Dehaene, Bossini, & Giraux, 1993). This finding lead to the idea that analog numerical magnitudes are represented via a positional coding along a spatial continuum in which numerical magnitudes are assumed to be mapped onto mental space from left to right in ascending order.(e.g., Daar & Pratt, 2008; Fias, 2001; Dehaene et al., 1993). This is the mental number line hypothesis (MNL). The SNARC effect was first described by Dehaene and colleagues (1993), and has been observed and investigated in multiple studies since (for reviews: Fias, & Fischer, 2005; Hubbard, Piazza, Pinel, & Dehaene, 2005; Wood, Willmes, Nuerk, & Fischer, 2008). Interestingly, the SNARC effect can be observed in tasks that do not require encoding the magnitude of the numbers presented, ordinal processing was not part of the requirements of the parity judgment task of the original experiment. This has led researchers to think of the SNARC effect as an automatic association between numbers and space. The mental number line is a useful metaphor to capture the spatial coding of numbers, however it must not be taken literary. There is no evidence for topographic organization of number-selective neurons (Nieder, Friedman, & Miller, 2003; Verguts & fias, 2004). Rather, because the spatial association of numbers are highly task-dependent, a careful position would consider those associations as part of our strategic use of knowledge, and as a result (Fias, & Fischer, 2005, p. 52). Moreover, evidences of the flexibility of spatial associations challenges the appropriateness of the number line metaphor. The existence of vertical as well as horizontal SNARC (Schwarz & Keus, 2003) and the systematic association of odd numbers with left space and even numbers with right space (Nuerk, Iversen, & Willmes, 2004). Further research will have to determine the extent to which the wide range of spatial numerical association can help as understanding the strategic nature of the cognitive representation of numbers.

Nonverbal counting would be pointless without the ability to arithmetically manipulate mental magnitudes so generated. Barth (2001) tested adults humans on addition, subtraction, multiplication and division of non-verbally estimated numerosities (dot arrays or tone sequence), subjects were presented with two numerosities in rapid sequence, each presentation too quickly to be verbally countable. Then, they were presented with a third numerosity. Subjects indicated if the result of the first two numerosities was greater or smaller than the third by pressing one of two buttons. Bath's experiments establish by direct test the human ability to manipulate non-verbal estimates of numerosity in accord with the prescribed arithmetic operation. Moreover, the accuracy of the comparisons was inversely proportional to the ration between the magnitudes to be compared. This result suggests that the scalar variability found in the nonverbal estimates of numerosity extend to the mental arithmetic operated with such magnitudes. Moreover, in a study by McCrink, Dehaene, and Deahene-Lambertz (2007) it has been argued that also the spatial representations of numerical magnitude extend to the domain of mental arithmetic. The authors found a cognitive bias in numerical estimations after mental calculation for the processing of non-symbolic numerosities. Participants viewed moving dot patterns being added or subtracted from one another and indicated whether the numerosity of a final set of dots was correct or incorrect. Surprisingly, in the case of addition, the subjects' estimated outcomes tended to be larger than the actual outcomes, whereas the estimations tended to be smaller than the actual outcomes with subtraction. McCrink and colleagues (2007) used the metaphor of a mental number line for analogue magnitude representations (Dehaene, 1997) to explain their finding. They speculated that mental calculations are functionally equivalent to movements along the spatial-numerical continuum and assumed that the overestimation after addition and the underestimation after subtraction reflect the subjects' tendency to move “too far” to either the right or left side. Since the observed effects are reminiscent of a perceptual phenomenon called representational momentum (Freyd & Finke, 1984), which represents the tendency of subjects to misjudge the stopping point of a moving object, McCrink and colleagues (2007) labelled the observed judgement bias after mental calculation the *operational momentum* (OM) *effect*. This phenomenon will be referred to as *operational bias* in order to disentangle the name from the specific space-related explanation.

Interestingly, a recent study of Pinhas and Fischer (2008) demonstrated that spatial response biases also emerge after mental calculations with exact numbers and provided thus first direct empirical evidence that the operational bias generalizes to symbolic arithmetic. Participants viewed addition and subtraction problems with Arabic digits and indicated the result by pointing to corresponding locations on a visually presented line that represented the numerical interval from 0 to 10 (see Siegler & Opfer, 2003, for a similar method). The analysis of the pointing end locations revealed that motor responses were systematically biased to the left side after subtracting and to the right side after adding. The finding of an operational bias for number processing has been interpreted as evidence that each approximate mental calculation, even when the input magnitude information is presented symbolically as Arabic numeral, relies on the same analogue magnitude code as the processing of non-symbolic numerosity information.

Common magnitude system

Resuming, the combined efforts of many researchers are advancing our understanding of how number is represented. Researchers studying numerical reasoning in adult humans, developing humans and non-human animals are using a suite of behavioral and neurobiological methods to identify the cognitive format and neural substrates of numbers and numbers manipulation. The image emerging from this effort is that adult humans share with non-human animals a system for representing number as language-independent mental magnitudes and that this system emerges early in development. The foundations of these mathematical abilities were probably present early in our evolutionary history and can be seen early in human development. Similarly, although adult humans use language to exceed the precision of this phylogenetically old system, they nevertheless simultaneously possess a phylogenetically and developmentally conservative system for representing number without language. This representational system, usually referred to as common magnitude system (CMS), would codes for quantity across modalities, with the mental representation proportional to the magnitude being represented. This system treats discrete quantities (e.g., three items) as analogous to continuous magnitudes, and due to increasing variability as the quantity represented increases, this system operates as a function of Weber's law. As the ratio of two magnitudes approaches 1:1 they become harder to discriminate, and beyond a certain threshold determined by the subject's 'Weber constant' they cannot be discriminated at all (e.g., Brannon, & Terrace, 1998; Brannon, & Terrace, 2000; Cantlon, &

Brannon, 2006; Halberda, & Feigenson, 2008; Jordan, & Brannon, 2006; Church, & Meck, 1984; Xu,& Spelke, 2000). The mental number line hypothesis is an alternative to—but not necessarily in contrast with—the idea of analog numerical magnitudes sharing a common representation with other magnitudes, like time, space, and sensorimotor magnitudes since different kinds of magnitudes in many cases must be combined to drive important behavioral decisions (Dehaene, 1997; Gallistell & Gellman, 2000; Brannon & Roitman, 2003; Walsh, 2003; Lourenco & Longo, 2011).

Evidence for AMS comes from a growing number studies reporting *within magnitude interferences*, i.e. interactions between numerical magnitude and magnitudes in sensorimotor control (for a review, see Bueti, & Walsh, 2009; Lindemann, Abolafia, Girardi, & Bekkering, 2007; Andres, Davare, Pesenti, Olivier, & Seron, 2004). Here is presented a short review on some example of within magnitude interferences, the relation between number and space have been already shown so it will be not further mentioned here.

Number and Time. Recent experiments demonstrated a connection between number magnitude and time (e.g., Vierck & Kiesel, 2010; Xuan, Zhang, He, & Chen, 2007). For example, Xuan et al. (2007) Asked participants to judge which of two successively presented stimuli displays were shown for a longer duration. In one of their experiments, the displays contained irrelevant digits that were either small or large and that could be displayed for short or long durations. If digit values were congruent with the display duration, i.e. small digits with short durations or large digits with long durations, fewer errors were made. This experiment clearly established a connection between digit magnitude and time on perceptual processing.

Number and Intensity. There are only few studies that points to a connection between number magnitude and intensity up until now (Vierck and Kiesel 2010; Lindemann, Abolafia, Girardi, & Bekkering, 2007). For istance, Vierck and Kiesel (2010) used a parity judgment task and asked participants to press a force-sensitive key weakly or forcefully for odd or even numbers. They found a congruency effect between the mode of response (weak/forceful) and the number magnitude (small/large), that is, when weak responses were related to small numbers and forceful responses were associated with large numbers, responses were faster and error rates smaller. The above findings provide indirect evidence for a connection between number magnitude and intensity.

In light of the discussed evidences it seems to me unnecessary to invoke a specifically spatial representation of numbers, as claimed by the mental number line hypothesis, because space and number draw upon common magnitude mechanisms. Nether the less those two hypothesis are not mutually exclusive because an abstract magnitude representation may relate on different specific representations depending on the specific task demands, including a spatial one. Thus, in order to provide direct evidences for this general magnitude account of number representation, a series of experimental works have been designed.

Chapter 2

Three experimental studies

Study 1: A Functional Measurement Study on Numerosity

The way that approximate numerical magnitudes are manipulated in order to judge (as opposed to calculate) the result of an arithmetic operation can be conceptualized as a multi-attribute judgment, with which the result is derived from the integration of the operands with a specific integration rule. This conceptualization allows for the application of the tools of Information Integration Theory (IIT) (Anderson, 1981; Anderson, 1982) to the study of mental arithmetic. Busemeyer (1991) summarizes some of the applications of IIT to the problem of intuitive estimations of algebraic operations on symbolic quantities (numbers) and continuous quantities (line lengths, tones, or weights). Moreover, in the field of IIT, many works use functional measurements to assess numerosity (Cuneo, 1982; Shanteau, Pringle & Andrews, 2007). Interestingly, at my knowledge, no study has yet applied this approach to the investigation of the way in which the results of arithmetic operations with discrete quantities are computed or approximated. Thus, in the present work, the applicability of the IIT approach to arithmetic of mental quantities was tested through the evaluation of the shape of the response function and of the goodness of fit of the model to behavioral data. Two different response methods have been used to support the generality of the result and to confront their peculiar features. Participants judged the average numerosity between two sequentially presented dot patterns to perform an approximate arithmetic task. In a first experiment, the response was given on a 0–

20 numerical scale, a well known method called categorical scaling, whereas in a second experiment, the response was given by the production of a dot pattern the numerosity of which was controlled by the rotation of a knob a relatively new method called numerosity production (Lindemann & Tira, 2011). The matching results of the two condition would validate the *numerosity production response method* as a viable method to asses mental magnitudes.

Functional measurment

IIT describes the psychological processes underlying multi-attributes decision-making and proposes a general method that is applicable to several contexts. IIT proposes a theoretical framework (cognitive algebra), which is accompanied by a methodology (functional measurement) that is relevant to the evaluation of its adaptation to the real contexts of the proposed models. IIT conceives the cognitive processes that lead to the integration of more information in a single concept (from physical stimuli S to a behavioral response R), as divided into three phases: evaluation, integration, and response. Each of these phases is governed by a specific function (s=V(S), r=I(s1,s2, …, sn), R=M(r)). This evaluation process leads to the assignment of an implicit value s to the individual constituent parts of the stimulus S. This is followed by an integration of these values that, in turn, leads to the formulation of an overall judgment. At this level, the different models that describe the ways the operation of integration is performed play a crucial role. The cognitive algebra framework provides three models of the integration process: the additive model (Anderson, 1962), the multiplicative model (Anderson & Shanteau, 1970; Anderson & Weiss, 1971), and the weighted average model (Anderson, 1965; Norman, 1976), also known as averaging. Thus, from this perspective, an algebraic operation can be considered as a process of the evaluation of a stimulus S, in which the operands are the constituent parts (S1, S2) of that stimulus. From this point of view, the process of evaluation includes the assignment of internal and subjective values, s1 and s2 to S1 and S2. This is followed by an integration of these internal values with an integration function. This leads to the formulation of an overall judgment, which represents the result of the algebraic operation. The functional measurement theory includes, besides each s value, a weight parameter. The weight represents the importance, assumed by the particular attribute in the overall judgment, and it is indicated by the parameter w in the models. Despite the fact that the theoretical formulation implies a distinction between scale values and weights, in both the additive and multiplicative models, the two parameters are not really distinguishable (Anderson, 1981). The effect of each attribute cannot be separated into a scale value and a weight. Conversely, the averaging model has the capability, under specific conditions, to distinguish between scale values and weights (Zalinski & Anderson, 1989). The averaging model of IIT represents the subject's response to a multi-stimulus situation, as a weighted average. Each stimulus has two parameters: the weight w, which conveys the importance of the stimulus on the final judgment, and the scale value s, which represents its position on the dimension of response (Zalinski & Anderson, 1991). The averaging model represents the integrated response, r, as:

$$r=\frac{\sum w_t s_t}{\sum w_t}, \quad t=1,2,\ldots \tag{1}$$

whereas, in a two stimuli situation, becomes:

$$r=\frac{w_1 s_1+w_2 s_2}{w_2+w_2} \tag{2}$$

The weight-value representations are common, but they are arbitrary in most formulations. Each weight in a standard regression model, for example, is confounded with the unit of the scale. The averaging model makes weight mathematically identifiable, and the empirical success of the model makes it psychologically meaningful (Zalinski & Anderson, 1991). The averaging model assigns weight and scale values to each stimulus. If all of the levels of one factor have the same weights wAi = wA, then the model is said to be equally weighted; if at least one of the levels differs, then the model is said to be differently weighted. Functional measurement makes use of the joint manipulation of at least two factors, according to a factorial design; the second block of each experiment was carried out for this purpose. From now on, it will be refer to as the factorial design block. Moreover, to differentiate the averaging model from the additive and multiplicative models, one or more factors at a time must be excluded from the factorial design; this is called a sub-design. The first experimental block of each experiment was meant explicitly for this purpose. From now on, the first block will be referred to as a sub-design block.

Experiment 1: categorical scaling

The aim of Experiment 1 was to study the integration rule, involved in approximate averaging operations of discrete quantities, and to evaluate the goodness of the fit of the averaging model to the data. First, dot sets were presented to participants with the instruction to indicate the numerosity of the sets on a 0–20 numerical scale. Later, the participants were asked to indicate on the same 0–20 scale the average numerosity between two sequentially presented dot sets. To test the integration rule that was involved in the task, the number of presented dots was systematically varied in a factorial design and in sub-designs. If the participants responded on a linearly distributed scale, and if they used an averaging integration rule to evaluate the averaging numerosity, then the plot of the complete factorial design was expected to be a bundle of parallel lines, along with lines that represent sub-designs, intersecting the bundle.

Method

Participants. Fifteen undergraduate female students from the University of Padua participated in the experiment. The average age of participants was 21.5 years (SD = .5). A convenience sampling was used, and the participants received no payment.

Apparatus. Participants used a keyboard and a computer screen in a quiet room. The distance between the subject and screen was 70 cm. A Python program was developed in order to process the input from the keyboard and to control the presentation of stimuli. Importantly, the spatial pattern of the appearance of the dots was unpredictable. Precisely, with every .6 degree of clockwise rotation, one additional dot (2 mm in diameter, .16° of visual angle) was presented at a randomly chosen free position within an unmarked circular target area of 140 mm in diameter (11.42° of visual angle), centered on the screen. The minimum distance between the two dots was .25 mm (.02° of visual angle).

Materials. The random dot patterns were presented in white on a black background. A circular gray area with a radius of 140 mm was presented to the participants just before the dot pattern, as an attention clue. Patterns of 0, 17, 38, 60, or 82 dots composed the presented stimuli; with an exception of the zero, the number sequence is a geometric series on a logarithmic scale. Stimuli consisted in white dots displayed in random positions in order to prevent the constitution of patterns that may have otherwise influenced the results. Random patterns are usually considered as preferable to other configurations because the perceptual structures of the dot patterns could affect their apparent visual numbers (Frith & Frith, 1972; Ginsburg, 1976; Krueger, 1972). A circular area with a fixed radius was used in order to prevent the number of dots from being proportional to the occupied area. A similar configuration has been widely used in many other experiments on this topic such as the studies by Knops, Viarougue, & Dehaene (2009) and Piazza, Izard, Pinel, Le Bihan, & Dehaene, (2004).

Procedure and Design. Participants were required to rate the numerosity of the presented dot patterns on a 0–20 numerical scale (Anderson, 1962). Participants were instructed to consider the response scale with none (zero) and very many (20) as scale ends. Participants were also instructed to type the numerical scale point value that they rated on a keyboard. Each subject was shown three blocks: the training block and the sub-design block; for which the subjects were asked to rate the numerosity of sets of dots; and the factorial design block, for which the subjects were asked to rate the average numerosity between two sequentially presented dots sets (the experimental procedure is depicted in Figure 1). Participants were instructed to respond as quickly and accurately as possible and to not to try to count the dots.

Each trial was composed of a presentation part and a production part. In each presentation part, a circular gray area was shown at the center of the screen for 1000 ms, followed by the presentation of a dot pattern for 2000 ms. This gray area/dot pattern sequence was repeated twice. At the end of the presentation part of the trial, a hash mark (#) was presented for 1000 ms. The disappearance of the hash mark indicated the beginning of the response phase, in which the participants could type their responses. Participants typed their responses on a field on the screen by typing on a keyboard. After the participants made their judgments, they pressed abutton to move on to the next trial, which started after an inter-trial interval of 500 ms.

Two subjects were excluded from data analyses because they did not show any response consistency. Training block. Eleven trials were administered in order to familiarize the participants with the specific task and response method before the experimental blocks were given. Unlike the experimental blocks, in the training blocks, only one quantity per trial was shown and feedback for the participants was provided after each trial. The training block provided stimuli with a number of dots that ranged from 0–100, which represents the two anchors of the scale (Anderson, 1982). As a form of feedback, the computer provided the closest value on the 0–20 scale to the number of shown dots, divided by 5. This training allowed for the calibration of the judgments of numerosity and minimized the variability, caused by inter-individual differences in the perception of non-symbolic numerosity (see Izard & Dehaene, 2008).

Sub-designs block. Two dot patterns were presented. The participants were asked to rate the numerosity of one of them, either the first or the second, as indicated by a signal (number 1 or 2), presented after the disappearance of the second dot pattern. Each pattern could have one of five different numbers of dots: 0, 17, 38, 60, and 82. This five-by-five design yielded 25 pattern pairs. However, because no judgment different from 0 is plausible or informative, in response to “an empty” screen as a stimulus, target patterns with 0 dots were omitted; accordingly, only 20 (i.e., 4x5) pattern pairs were presented. Each pattern pair was presented twice, and each time, the pattern pair was presented with a different indication of the pattern to rate (number 1 or 2) for a total of 40 trials, presented in randomized order. In summary, five responses were collected for each dot pattern to be evaluated, the mean of the five responses was used in the following statistical analysis.

Factorial design block. Participants had to rate the average quantity of dots between two presented patterns. Each pattern could have one of five numbers of dots: 0, 17, 38, 60 and 82. This five-by-five design yielded 25 pattern pairs. However, because no judgment different from 0 is plausible for pattern pairs with 0 dots, the (0, 0) pair was not presented; accordingly, only 24 (i.e., 5x5-1) pattern pairs were presented. Each pattern pair was presented 5 times for a total of 120 trials, presented in randomized order. In summary, five responses were collected for each pair of dot patterns to be evaluated, and the mean of the five responses was used in the following statistical analysis. Each complete session of the experiment lasted approximately 30 minutes. Before every block, instructions were printed on the screen. Participants were requested to read the instructions and explain them back to the experimenter to verify that they understood correctly.

Results

Psychophysical function. The shape of the response function of the sub-design block was tested. The shape of the response function using a magnitude estimation response methods is generally best described by a power function, $R = \alpha \times n^{\beta}$ with an exponent β smaller than 1 (Izard & Dehaene, 2008; see also Siegler & Opfer, 2003). In order to test the shape of the response function of the numerosity production response method, a logarithmic regression analysis wasperformed (see for instance Seber & Wild, 2003) for the estimations of each numerosity, averaged across subjects. Remarkably, the regression of the averaged data fits very well in $r^2 = .76$, and the resulting response function was $y = .93 \times n^{.73}$.

Model identification. The responses were analyzed in order to assess the plausibility of integration models. The classic approach, used by the functional measurement for the individuation of the integration function of the model, is the analysis of variance (ANOVA). The theorem of parallelism (Anderson, 1981) argues that if the integration model is additive, the graph of marginal means will appear as a bundle of parallel lines. Morever, any observed deviation from the parallelism will be purely due to the component of error. Thus, an ANOVA was conducted. Because of the interaction between the two factors (1st and 2nd dot pattern) was not significant ($F(15,14) = 1.52$, $p = .08$), the deviation from parallelism can be considered negligible, and the multiplicative model can be discarded from the candidates (see Figure 2). Moreover, a significant main effect was found for both factors: 1st dot pattern ($F(4,14) = 198.45$, $p < .001$, $\eta_p^2 = .309$) and 2nd dot pattern ($F(4,14) = 254.52$, $p < .001$, $\eta_p^2 = .36$). The test of the opposite effects (Anderson, 1981) is used to distinguish an additive model from an averaging one. This test makes use of the methodology of the sub-designs (Norman, 1976; Anderson, 1982). This methodology consists of associating the full factorial design with one or more sub-design(s) that exclude(s) one or more factor(s) at a time; the first experimental block was created explicitly for this purpose. The two factors (1st and 2nd dot pattern) were modified, adding to each one a level based on the responses of the sub-designs, referred to in the 1st and 2nd dot pattern. If the model was not additive but averaging, then a significant interaction of the two factors was expected. Indeed, the ANOVA showed a significant interaction effect ($F(15,14) = 2.30$, $p < .001$, $\eta_p^2 = .014$). Then, the parallelism observed in the full factorial design, along with the significance of the interaction, obtained when the sub-designs were added, might be considered as evidence in favor of the averaging model with equal weights within factors. It is the so-called equal-weight averaging model (EAM)(Wang & Yang, 1998). Moreover, for every factor, the significance of the main effect was found to be practically unaffected by the introduction of the new level, 1st dot pattern ($F(4,14) = 199.00$, $p < .001$, $\eta_p^2 = .251$), 2nd dot pattern ($F(4,14) = 249.52$, $p < .001$, $\eta_p^2 = .296$).

Model estimation. After the model was identified, the averaging model parameters for each participant were estimated with the R-average method (Vidotto & Vicentini, 2007; Vidotto, Massidda, & Noventa, 2010) and the implemented R-average package, version 0.4-0. The following analyses were computed on the estimated model parameters of all the participants, except when noted. The adaptation of the models to the data was evaluated, in terms of the adjusted r^2 for each subject, showing that the model fit the data very well for all of the participants of Experiment 1 with median $r^2_{adj} = .84$ (ranging between .78 and .99). As previously mentioned, the differential-weights model (DAM) was rejected, due to the lack of significant effects in the interaction between the linear components of the factors (Anderson, 1982). The EAM weights of the 1st and 2nd dot patterns were compared1, revealing no significant difference ($t(14) = -.60$, $p = .21$). Under a principle of parsimony, this notion led us to opt for an averaging model with equal weights between factors (wA

= wB), which was called a simple averaging model (SAM). A generalized linear mixed model (GLMM) was then applied to the s parameters of the SAM model, using the participants as random variables and the two factors, numerosity (0, 17, 38, 60, 82) and dot pattern (1st, 2nd), as fixed variables. The results showed a significant effect of the factor numerosity ($\chi^2(4) = 1786.61$, $p < .001$) with a strong and significant linear component. No statistically significant difference was found in the main effect for the dot pattern ($\chi^2(1) = 2.82$, $p = .27$) or interaction between numerosity and dot pattern ($\chi^2(4) = 3.66$, $p = .17$), showing that the difference between the two dot patterns in the s parameters was negligible.

Worth noting is that the maximum level of uniqueness (for w) is a common ratio scale. The unit of this scale is arbitrary because all the weights may be multiplied by

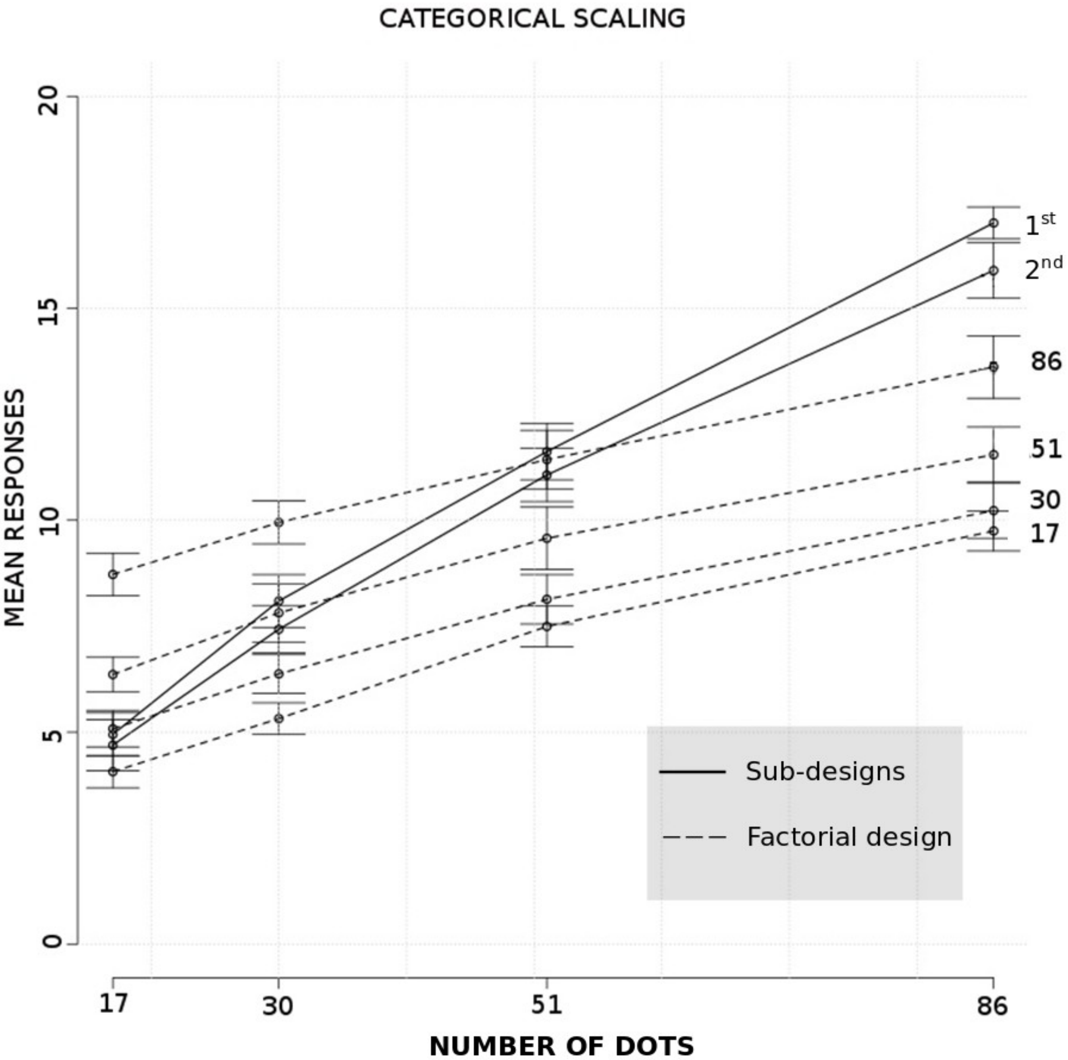

Figure 1: Experiment 1: plot of the subjects' estimations using categorical scaling method (mean responses are on the y-axis). In the complete factorial design, the number of dots identifies dashed lines for different numerosities of the 1st patterns while numbers of dots for the 2st pattern are in the x-axis. In the two sub-designs, 1st and 2nd identify continuous lines for the 1st and 2nd dot patterns while numbers of dots for the pattern are in the x-axis.

a constant without changing the model prediction" (Anderson 1982, p. 89). Now, considering log(w), the origin of scale is arbitrary but no more the unit, indeed all the log(w) may be added by a constant with no change in the model prediction (Vidotto, 2013). In such a way the mean of log(w) has the property to be reference invariant and the standard deviation of log(w) has the property to be absolutely invariant for any vertical translation; indeed, the t-test for differences was applied to log(w).

Response latencies. The average latency to perform a categorical scaling was 4177.63 ms with a standard deviation of 1969.01 ms. It is important to note that the latencies were not correlated with the number of dots (r = .04). This result ensures that the participants were not using counting strategies; otherwise, an increase in the reaction time with increasing numerosity would have been expected (Akin & Chase, 1978; Mandler & Shebo, 1982; Trick & Pylyshyn, 1993).

Experiment 2: numerosity production

The aim of Experiment 2 was to test the appropriateness of a new method of numerosity production to IIT studies. Participants were asked to indicate the numerosity of one presented dot pattern or the average numerosity between two sequentially presented dot patterns by producing that number of dots on the screen. Participants controlled the number of dots of their responses by turning a knob in a clockwise or counterclockwise direction. To test the appropriateness of the method, the number of presented dots was varied systematically in a factorial design and sub-design. As in the previous experiment, the integration rule was also studied and the goodness of fit of the averaging model to the data was evaluated. If participants responded on a linearly distributed scale, the plot of the complete factorial design would have been expected to form the shape of a bundle with parallel lines.

Method

Participants. Fourteen undergraduate female students from the University of Padua participated in the experiment. The average age of participants was 20.2 years (SD = .5). A convenience sampling was used. he participants received no payment.

Apparatus. The apparatus was identical to that in Experiment 1, except for the response device. The response device was a custom-made knob of 4.50 cm in diameter and 1.50 cm in height. The response device was also mounted on a small box (6 cm × 15 cm × 15 cm) and placed on a table. The knob was connected to a computer with a USB interface and could be rotated both clockwise and counter-clockwise. Knob rotation axis was parallel to the Cartesian z-axis. A Python program was developed to process the knob input and to control the stimulus presentation. The more the knob was rotated in a clockwise direction, the greater the number of dots that appeared on the screen. Rotation in the opposite direction decreased the number of dots, until no dots were left on the screen. Importantly, the spatial pattern of the appearance or disappearance of the dots was unpredictable. With every .6 degree of clockwise rotation, one additional dot (2 mm in diameter) was presented at a randomly chosen free position within an unmarked circular target area of 70 mm in diameter, centered on the screen. Counter-clockwise rotation deleted randomly selected dots from the display. The minimum distance between the two dots was .25 mm. The maximum number of dots was limited to 300.

Materials. The materials used in Experiment 2 were the same as those in Experiment 1.

Procedure and Design. The procedure and design were identical to those of Experiment 1, except for the response method that consisted of rotating the knob to produce the desired quantity of randomly distributed white dots. Thus, the response method in this experiment was a numerosity production, instead of a categorical scaling. At the beginning of the response phase, participants could rotate the knob in order to perform the numerosity production task. Participants always started the response phase with zero dots on the screen and turned the knob clockwise to increase the number of dots or counter-clockwise to decrease it.

Results

Psychophysical function. As for the data of the sub-design in Experiment 1, in order to test the shape of the response function of the categorical scaling response method, a logarithmic regression analysis was performed for the estimation of each numerosity, averaged across subjects. The regression of the averaged data fitted acceptably with $r^2 = .58$. The resulting response function was, $y = .15 \times n^{.58}$.

Model identification. In Experiment 2, the responses were analyzed in order to assess the plausibility of integration models by performing a 5x5 ANOVA with 1st and 2nd dot pattern as factors. Because the interaction between the two factors was not significant ($F(15,110) = .82$, $p = .64$), the deviation from parallelism can be considered negligible, and the multiplicative model can be discarded for this experiment, as it was for Experiment 1. Moreover, a significant main effect was found for every factor, 1st dot pattern (F(4,11) = 162.35, p < .001, η_p^2 = .28), 2nd dot pattern (F(4,11) = 227.90, p < .001, η_p^2 = .358). The two variables (1st and 2nd dot pattern) were modified, adding to each one a level that was made from the ratings on single-dot patterns (the sub-design). If the model was not additive, but averaging, it was expected that the addition of the new levels to the factors would involve a significant interaction between the two. Subsequently, the ANOVA showed a significant interaction effect (F(15,110) = 2.21, p < .001, η_p^2 = .01). Then, the parallelism between the factors of the full factorial design and the interaction, caused by the adding of sub-designs (see Figure 3) was found to be evidence in favor of EAM, as in Experiment 1. Moreover, for each factor, the significance of the main effect was found to be practically unaffected by the introduction of the new level, 1st dot pattern (F(4,11) = 172.12, p < .001, η_p^2 = .241), 2nd dot pattern (F(4,11) = 229.78, p < .001, η_p^2 = .298).

Model estimation. After the model was identified, the averaging model parameters for each participant were estimated with the same procedure that was previously applied in Experiment 1. The adaptation of the models to the data was evaluated in terms of adjusted r2, showing that the model fitted the data very well for almost all of the participants with median r2adj = .85 (ranging between 71 and 99). As previously mentioned, the DAM was rejected because it did not present a significant effect in the interaction between the linear components of the factors. The EAM weights of the first and the second dot patterns were compared, which revealed no significant difference (t(11) = 1.79, p = .56). This led us to opt for a SAM. A GLMM was then applied to the s parameters of the model, using the participants as random variables and the two factors of numerosity (0, 17, 38, 60, 82) and dot pattern (1st, 2nd), as fixed variables. The main effect for numerosity was found to be statistically significant ($\chi^2(4) = 4044.78$, $p < .001$) but that was not true for dot pattern ($\chi^2(1) = 1.50$, $p=.13$) or for the interaction between numerosity and dot pattern ($\chi^2(4) = 5.02$, $p = .10$). This shows that the difference between the two dot patterns in the s parameters was negligible.

Response latencies. The average latency to perform a categorical scaling was 3880.13 ms (SD = 2229.38). Importantly, the latencies were not correlated with the number of dots ($r = .039$), ensuring that the participants were not using counting strategies, which was also the case in Experiment 1.

Discussion and conclusions

In both experiments, the participants responded quickly, and their response times did not increase with numerosity. This reveals that the participants did not use counting strategies but instead, based their judgments on approximate numerosity estimation. In both of the experiments, the results of the analysis on the estimated averaging values seemed to indicate that the subjects' estimations are best described by an EAM. Moreover, the weights of the two dot patterns do not appear to differ significantly, suggesting the use of a SAM. Accordingly, the scale values vary, depending only on the numerosity of the stimulus and are unaffected by its position (1st or 2nd dot pattern). This demonstrates that neither the effect of primacy nor the effect of recency influence the evaluation of the average numerosity, despite the sequential temporal order of the presentation of the stimuli (Busemeyer, 1991). In other words, this means that the participants give the same importance to the two quantities of each trial during averaging operations. In both experiments, the adjusted r^2 showed that SAM was able to explain a very great portion of variance for almost all of the participants; this supports the explanatory capability of the averaging model, applied to mental arithmetic problems with discrete quantities. Since the participants were instructed to perform an averaging operation, the factorial plot should exhibit parallelism, if the response measure was on a linear scale. As shown in Figures 2 and 3, and according to the results of the full factorial design ANOVA (without sub-designs), the rating data (Figure 1) and the numerosity production data (Figure 2) show clear parallelism. This allows researchers to validate the numerosity production, as a response measure on a linear scale, a prerequisite for a method to study stimulus interaction, and for the analysis of non-linear integration rules (Anderson, 1982).

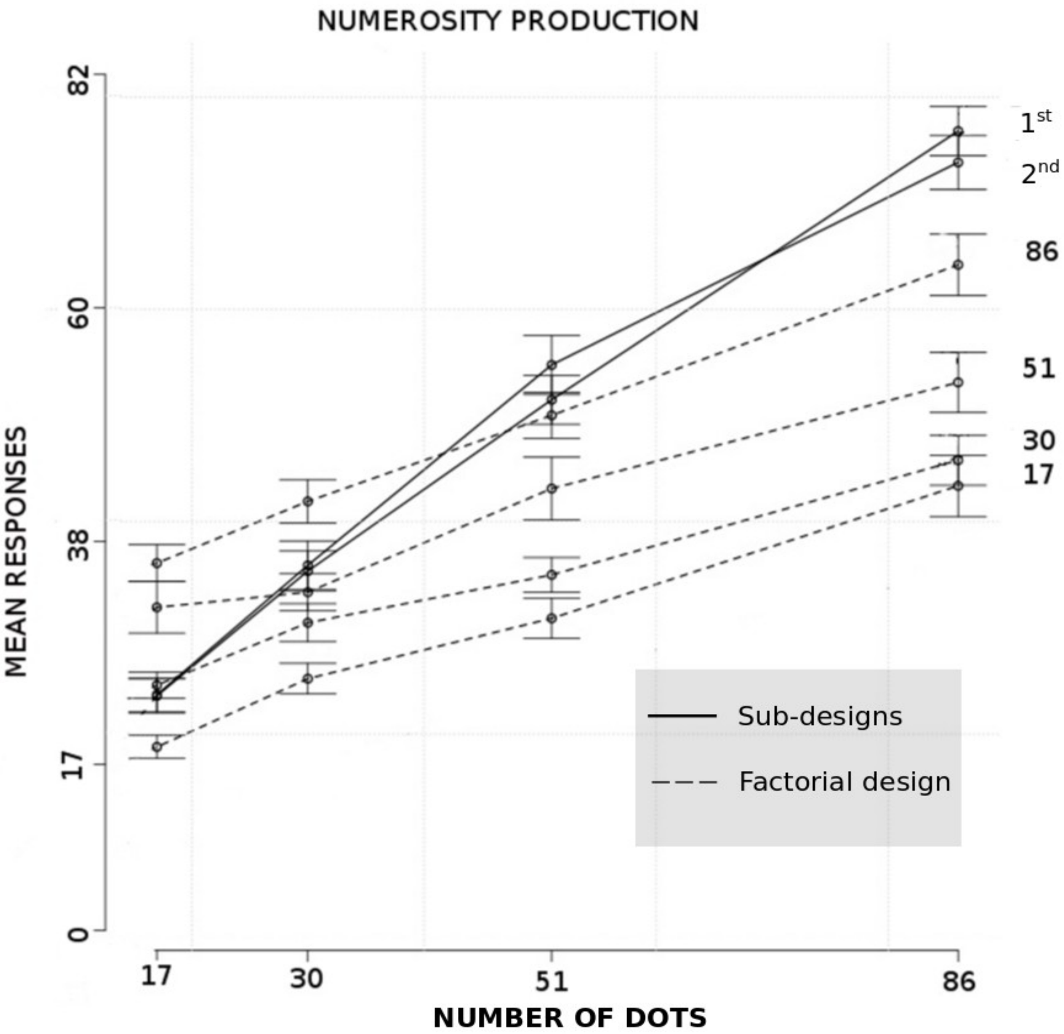

Figure 2: Experiment 2: plot of the subjects' estimations using numerosity production method (mean responses are on the y-axis). In the complete factorial design, the number of dots identifies dashed lines for different numerosities of the1st pattern while the numbers of dots for the 2nd pattern are in the x-axis. In the two sub-designs, 1st and 2nd identify continuous lines for the 1st and 2nd dot patterns while the numbers of dots for the pattern are in the x-axis.

The linearity of the response scale and the similar trends of the two response methods paves the way for further interesting possibilities of application of the IIT framework for the numerosity production response method. The applicability of IIT to mental arithmetic problems with discrete quantities is supported with the following factors: the linearity of the scale observed with both of the response methods and the high explanatory capability demonstrated by the averaging model in both experiments. On the other hand, since a series of stimuli were used, composed by dot collectivities, distributed on a fixed radius circular area and manipulated the number of dots, it may be argued that the density of dots in each stimulus may have influenced the participants' impressions of numerosity (Krueger, 1972; Allik & Tuulmets, 1991; Shanteau et al., 2007). We believe that the variation in density, along with the levels of the factors, do not weaken our conclusions. This is because even if the numerosity judgment was based on the density of the stimulus, it does not change the way that the internal representations of the stimuli were integrated. Furthermore, this does not change the conclusions about the parallelism and linearity of the response functions. Since the effect of over- or under-evaluation, linked to the specific density of each level of each factor is proportional to the size of the stimulus, and since it remains constant for that level to every proposition in the factorial design, this does not affect the nature of the model but affects only its scale values.

The averaging model of IIT was established as a viable instrument in assessing mental arithmetic with discrete quantities; it is able to properly describe behavioral data, distinguishing between the value of the evaluation of a stimulus and its importance in the integration process. Moreover, a new numerosity production method was tested for the linearity of its response scale. Finally, averaging operations with discrete quantities appear to not be affected by the presentation order of the dot patterns. For all of these reasons, the IIT framework seems to be a promising approach, particularly for future applications in the field of mental arithmetic with discrete quantities.

e factorial

Study 2: Influence of Response Force on the Operational Bias Effect

It has been argued that our ability to learn and engage in highly complex mathematical thinking is ultimately grounded in inborn mechanisms to represent numerical quantities as analog magnitudes (Dehaene, 2004). Already, pre-verbal infants are able to extract numerical magnitude from sensory input and perform rudimentary arithmetic computations on them (Wynn, 1990; Xu & Spelke, 2000). Also, later in life, a tight connection between symbolic numerical skills and the analog number code prevails. For instance, adults take increasingly more time to decide which of two digits is larger as the absolute difference between the digits gets smaller. This so-called numerical distance effect suggests that the process of comparing two digits is similar to comparing two physical (analog) magnitudes (Moyer, & Landauer, 1967).Several authors have claimed that analog numerical magnitude shares a common representation with other magnitudes, like time, space, and sensorimotor magnitudes since different kinds of magnitudes in many cases must be combined to drive important behavioral decisions (Dehaene, 1997; Gallistell & Gellman, 2000; Brannon & Roitman, 2003; Walsh, 2003; Lourenco & Longo, 2011). For example, spatial information of actions can be used to make predictions about the immediate sensorimotor consequences of those actions (Rossetti & Pisella, 2002). This representational system—usually referred to as analog magnitude system (AMS)--codes for magnitude across modalities, with the mental representation proportional to the magnitude being represented. Evidence for AMS comes from a growing number studies reporting within magnitude interferences, i.e. interactions between numerical magnitude and magnitudes in sensorimotor control (for a review, see Bueti, & Walsh, 2009; Lindemann, Abolafia, Girardi, & Bekkering, 2007;

Andres, Davare, Pesenti, Olivier, & Seron, 2004). For instance, recent experiments demonstrated a connection between number magnitude and time (e.g. Xuan, Zhang, He, & Chen, 2007), reaction times and kinematics in grasping actions (Andres, Davare, Pesenti, Olivier, & Seron, 2004; Lindemann, Abolafia, Girardi, & Bekkering, 2007) and motor force (Vierck, & Kiesel, 2010). In addition, a common neural substrate has been reported for number, size, and luminance in the intraparietal sulcus (Pinel, Piazza, Le Bihan, & Dehaene, 2004).

Alternative to—but not necessarily in contrast with—the idea of a shared magnitude system, the mental number line hypothesis proposes that analog numerical magnitude is represented via a positional coding along a spatial continuum (e.g., Daar & Pratt, 2008; Fias, 2001; Dehaene et al., 1993). Numerical magnitude is assumed to be mapped onto mental space from left to right in ascending order. This idea is supported by a great amount of research showing systematic mappings between number and space (for a review, see Hubbard, Piazza, Pinel, Dehaene, 2005)—the most prominent demonstration of this mapping being the spatial-numerical association of response codes effect (SNAe factorialRC effect); (Dehaene, Bossini, & Giraux, 1993), which describes the finding of faster reaction times for small numbers when responses are made with the left compared to the right hand and the opposite pattern for large numbers.

Based on a recent study by McCrink, Dehaene, and Deahene-Lambertz (2007), it has been argued that spatial representations of numerical magnitude also extend to the domain of mental arithmetic. The authors found that participants systematically overestimate the results of addition operations and underestimate the results of subtraction –operations—an effect originally labeled operational momentum, which will be refered to as operational bias in order to disentangle the name from the specific space-related explanation of the phenomenon. The operational bias is usually interpreted as evidence for the idea that mental calculations are functionally equivalent to attentional shifts along the mental number line, and that the overestimation after addition and the underestimation after subtraction reflects the subject tendency to move “too far” to either the left or the right. However, it is important to note that the task used by McCrink and colleagues had no spatial component. In so far, the mental number line interpretation of the operational bias is, to a considerable degree, based on earlier findings in support of a spatial coding of numerical magnitude. We argue that it is unnecessary to invoke a specifically spatial representation of numbers to explain the operational bias. Next to the original interpretation of the operational bias, several alternative explanations have been provided (McCrink, & Wynn, 2009). According to one very parsimonious hypothesis, operational biases result from the use of the heuristic that "when adding, accept more" and "when subtracting, accept less".In principle, increasing and decreasing non-spatially represented analog numerical magnitudes could have the potential to elicit an over- and undershoot for additions and subtractions respectively. Therefore, we consider the idea that the operational bias is driven by processes within the shared magnitude system an alternative to the mental number line account.

In order to test the shared magnitude account of the operational bias, it was investigated whether required response force can modulate the operational bias. Participants were instructed to solve symbolic addition and subtraction problems and to generate a response by rotating a knob to produce the desired quantity of randomly distributed white dots. Crucially, by changing the resistance of the knob, the Required Force to turn the knob was manipulated. The Required Force to turn the knob has been varied in a three-level between-subject fashion (null, low, high). In the low and high force conditions, the knob was equipped with extension springs of increasing newton-millimeter rate; the further the knob was rotated in a clockwise direction, the more dots were presented on the screen and the more tense the spring, increasing the force required to turn the knob. In the null force condition, no spring was used and the knob was free to rotate. The operational bias effect has been assessed by comparing the response to addition and subtraction operations in the calculation task of each force condition.

It has been shown that mental calculations with carry-overs over the decade break are performed slower and are more error prone (Deschuyteneer, De Rammelaere, & Fias, 2005), resulting in more working memory load, as compared to multi-digit calculations without carry-overs (DeStefano & LeFevre, 2004; Imbo, Vandierendonck, & De Rammelaere, 2007). Since additions and subtractions with carry or borrowing involve apparently different cognitive processes than calculations without carry-over, the requirement to perform carry operations was systematically varied. Operations with the zero as the second operand were additionally included . These zero operations have been suggested to provide a measure of the "pure operational momentum effect" (Pinhas & Fischer, 2008) without activation of a second magnitude.

Based on the shared magnitude account of the operational bias, a modulation of the operational bias effect by Required Force was predicted. Based on the proposed coupling of numerical and sensorimotor magnitudes, the operational bias was expected to be positively correlated with Required Force.

Method

Participants. One hundred and five undergraduate students from the Radboud University Nijmegen (26 males) participated in the experiment in return for 7.5 euros or course credits. All reported having normal or corrected-to-normal vision. The average age of participants was 23.9 years (SD = 4.8). There were 35 participants randomly assigned to each of the three force conditions.

Apparatus. A keyboard, a computer screen, and a custom-made knob of 5.20 cm in diameter and 2.80 cm in height mounted on a small box (9 cm × 15 cm × 6.50 cm) placed in a quiet room were used. The knob was connected to a computer via a USB interface, and its rotation axis was parallel to the Cartesian z-axis. In two of the three experimental conditions, two different springs inside the device ensured that the knob switched back to its initial position when the participant let go of the knob (low force= 0.72 n/mm, high force=1.4 n/mm). A Python program was developed using the Expyriment package (Krause & Lindemann, 2013), to process the knob input and control the stimulus presentation. The more the knob was rotated in a clockwise direction, the greater the number of dots that appeared on the screen. Rotation in the opposite direction decreased the number of dots until none were left on the screen. Importantly, the spatial pattern of the appearance or disappearance of the dots was unpredictable. The distance between subject and screen was 70 cm. Precisely, with every .6 degree of clockwise rotation, one additional dot (2 mm in diameter, .16° of visual angle) was presented at a randomly chosen free position within an unmarked circular target area of 140 mm in diameter (11.42° of visual angle), centered on the screen center. The minimum distance between two dots was .25 mm (.02° of visual angle). The maximum number of dots was limited to 300. Each production phase started with an empty screen (i.e. zero dots).

Materials. The random dot patterns were presented in a white color on a black background, and numerals were presented in a white sans-serif font (height: 4 mm, width: 3-7 mm). All Arabic two-digit numbers served as targets for the number task (first block). For the calculation task (second block), a list of 24 addition and 24 subtraction operations was compiled (see Table 1).

Table 1: Addition and Subtraction operations used in the mental calculation task. The results were not presented during the experiment.

Addition operations	Subtraction operation	Addiction Zero operation	Subtraction Zero operation	Results
13 + 21 (n)	57 – 23 (n)	34 + 0	34 – 0	34
21 +14 (n)	69 – 34 (n)	35 + 0	35 – 0	35
12 + 24 (n)	98 – 62 (n)	36 + 0	36 – 0	36
14 + 23 (n)	58 – 21 (n)	37 + 0	37 – 0	37
13 + 29 (c)	91 – 49 (c)	42 + 0	42 – 0	42
29 + 14 (c)	61 – 18 (c)	43 + 0	43 – 0	43
18 + 29 (c)	74 – 27 (c)	47 + 0	47 – 0	47
12 + 41 (n)	74 – 21 (n)	53 + 0	53 – 0	53
18 + 36 (c)	72 – 18 (c)	54 + 0	54 – 0	54
24 + 32 (n)	68 – 12 (n)	56 + 0	56 – 0	56
19 + 38 (c)	71 – 14 (c)	57 + 0	57 – 0	57
41 + 17 (n)	89 – 31 (n)	58 + 0	58 – 0	58
19 + 43 (c)	81 – 19 (c)	62 + 0	62 – 0	62
14 + 49 (c)	82 – 19 (c)	63 + 0	63 – 0	63
16 + 48 (c)	82 – 18 (c)	64 + 0	64 – 0	64
53 + 14 (n)	79 – 12 (n)	67 + 0	67 – 0	67
16 + 57 (c)	92 – 19 (c)	73 + 0	73 – 0	73
28 + 46 (c)	92 – 18 (c)	74 + 0	74 – 0	74
47 + 28 (c)	91 – 16 (c)	75 + 0	75 – 0	75

29 + 47 (c)	93 – 17 (c)	76 + 0	76 – 0	76
51 + 32 (n)	97 – 14 (n)	83 + 0	83 – 0	83
63 + 21 (n)	97 – 13 (n)	84 + 0	84 – 0	84
12 + 73 (n)	98 – 13 (n)	85 + 0	85 – 0	85
52 + 34 (n)	98 – 12 (n)	86 + 0	86 – 0	86
Letters in parentheses indicate the Calculation type, c:carry operations, n: no-carry operations.				

We used stimuli with random positions to prevent the constitution of patterns which may influence the results. Random patterns are usually considered a reference in respect to other configurations to evaluate their effect and increase or decrease their apparent visual number (Frith & Frith, 1972; Ginsburg, 1976; Krueger, 1972). a circular area with a fixed radius was used to prevent the number of dots from being proportional to the occupied area. A similar configuration was widely used in many experiments (e.g., Knops, Viarougue, &configurations to evaluate their effect and increase or decrease their apparent visual number (Frith & Frith, 1972; Ginsburg, 1976; Krueger, 1972). A circular area with a fixed radius was used to prevent the number of dots from being proportional to the occupied area. A similar configuration was widely used in many experiments (e.g., Knops, Viarougue, & Dehaene, 2009; Piazza, Izard, Pinel, LeBihan, & Dehaene, 2004).

Procedure and Design

Numerosity production response method. Participants were required to produce a random dot pattern that corresponded to a previously presented target number, or the result of an operation, by rotating the knob with their right hand. If participants had the feeling that the numerosity of the dots was equivalent to the requested number, they finished their estimation with a key-press by their other hand. Each trial of each block started with a centrally presented attentional cue (‘#’ symbol), which was replaced after 500 ms by a two-digit target number, or an operation, depending on the block. As soon as the space bar was pressed, the target number, or operation, disappeared and the knob could be rotated. The numerosity productions always started from zero. Once participants made their judgment with a key-press and let go of the knob, the next trial started after an inter-trial-interval of 500 ms.Training block. Before the actual experiment started, participants were familiarized with the numerosity production response method in a short training session (20 randomly chosen trials from the number block stimuli set, always including 10 and 99). In contrast to the experimental session, written feedback about the amount of produced dots was provided after each trial. This training served as a calibration of numerosity judgments and minimized the variability caused by inter-individual differences in the perception of non-symbolic numerosities (see Izard & Dehaene, 2008).

Numerosity judgment block. Every participant performed first a numerosity judgment task in which they indicated the approximate magnitude of two-digit numbers. Each two-digit numeral (10 to 99) was presented once, resulting in a total of 90 trials. The order of trials was randomized. Mental calculation block. The task in the calculation block comprised 144 trials in total. Instead of a single number, an addition or subtraction problem was presented with the instructions for the participant to indicate the result of the presented problem as quickly and accurately as possible. A list of 24 addition and 24 subtraction operations has been compiled (see Table 1). Each problem comprised two operands. No decade numbers or symmetric numbers (e.g., 22, 33) occurred as operand or result. Operands and result of one problem never had the identical decade or unit digits. For additions and subtractions, half of the first and half of the second operands were odd numbers. The outcomes of addition and subtraction operations were matched. Half of the operations required a carry operation; the other halves were no-carry operations. Operands were chosen to approximately match the average results between the carry and the no-carry operations; indeed the mean result of the carry operations was 59.5 (SD=16.6) and 60.0 (SD=11.8) for the no-carry operations (see Table1). Furthermore, 48 zero operations were added by generating for each addition and subtraction problem a corresponding problem with the same result but with zero as the second operand. The trial order was randomized in both blocks. Each complete session of the experiment lasted for approximately 45 minutes. Before every block, instructions were printed on the screen. Participants were requested to read the instructions and explain them back to an experimenter as proof of their clear understanding.

Results

Response Latencies. The average latency to perform judgments was 3439.67 ms with a standard deviation of 2596.75 ms. Importantly, the latencies were not correlated with the number size (r = -.06). This result ensured that the participants were not using counting strategies; otherwise, an increase in the reaction time would have been expected with increasing numerosity (Akin and Chase, 1978; Mandler and Shebo, 1982; Trick and Pylyshyn, 1993).

Numerosity Judgment Task. If participants were capable of accurately judging numerosity, the number of produced dots should increase linearly with the target number. The shape of the response function using a magnitude estimation response method is generally best described by a power function, $y=\alpha n^{\beta}$, with an exponent β smaller than 1, that indicates a tendency to underestimation (Izard & Dehaene, 2008; see also Siegler & Opfer, 2003). In order to test the shape of the response function, an individual log-log regression was performed(see, for instance, Seber & Wild, 2003), regressing the log of the shown number against the log of the given judgment. Because a production response method was used instead of an estimation, the β is expected to be bigger than 1, indicating an overproduction of dots as found in a previous study (Lindemann & Tira 2011). The resulting average response function was y=-.03$n^{1.05}$; remarkably, the regressions fit the judgments of all participants nicely, median r^2 =.96 (ranging between r^2 =.88 and r^2 =.99). The slope of the regression function was significantly larger than one $t(104)=8.98$, $p<0.001$, $d=1.76$, indicating that participants, on average, produced too many dots. This numerosity overproduction, which increases with number size, is in line with the empirically well-established tendency of participants to underestimate perceived non-symbolic numerosities (e.g., Izard & Dehaene, 2008). Participants varied in their response functions with exponents ranging between 0.85 and 1.20. That is, some participants showed a logarithmically compressed relationship between asked and produced quantity, while others showed an expanded relationship.

Mental Calculation Task. Individual logarithmic regressions analyses were performed, regressing the log of the correct result of the proposed operations, against the log of the given judgment. The regressions fit the estimations very well for almost all participants, median r^2 =.91 (ranging between r^2 =.78 and r^2 =.99). The resulting average response function was y=-.06$n^{1.04}$. The slope of the regression function was significantly larger than one $t(104)=11.98$, $p<0.001$, $d=1.77$, indicating that participants, on average, produced too many dots. Again, some participants showed a logarithmically compressed relationship between asked and produced quantity, while others showed an expanded relationship. To analyze the variability of the estimations, the standard deviations of the estimations were individually regressed against the correct results. The slope of the regression function was significantly larger than zero $t(104)=27.00$, $p<0.001$, $d=5.29$, reflecting that the precision of judgments decreases with number size.

To investigate the effects of Required Force on mental arithmetic it was calculated, for each participant, the constant judgment error, defined as the average error between correct outcomes and responses (Schutz & Roy, 1973).

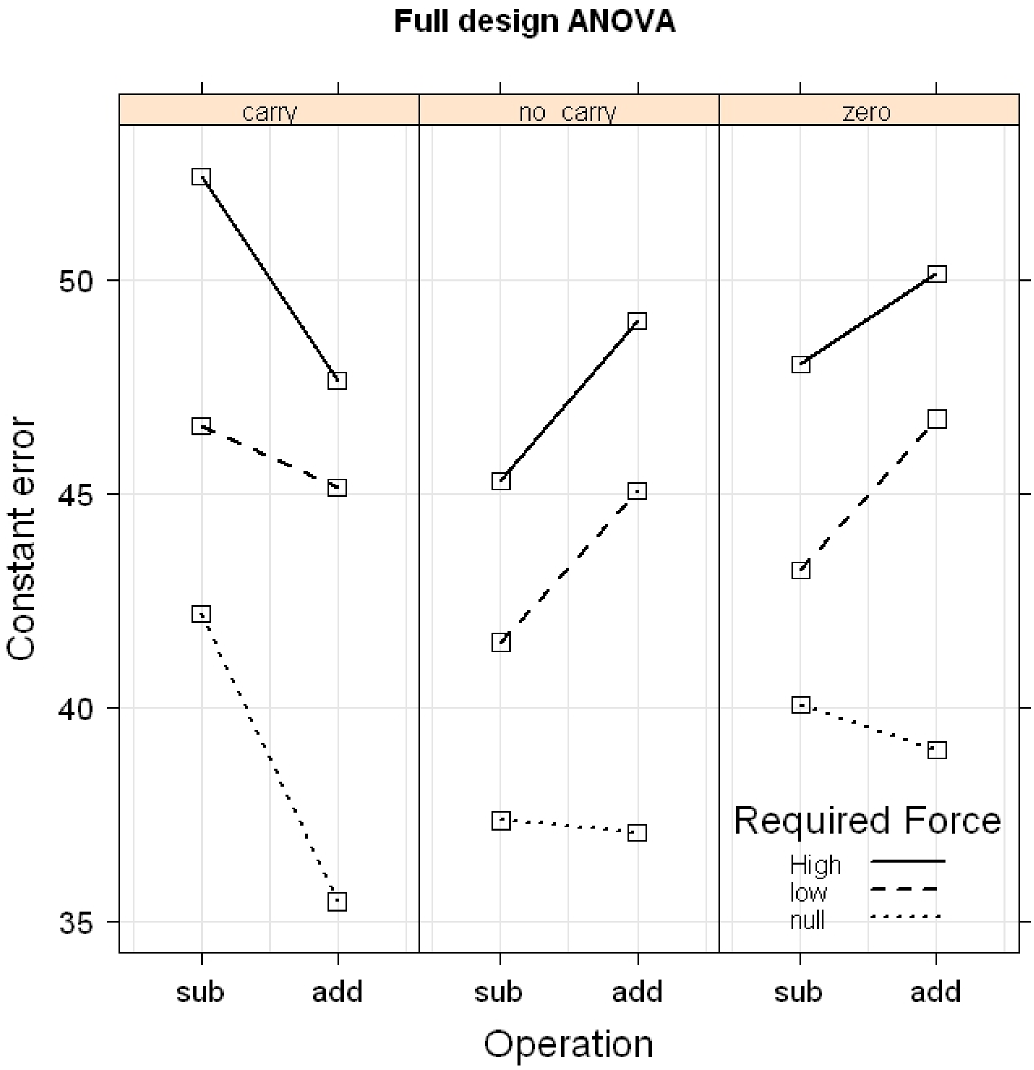

Figure 3: Mean constant error in the calculation task, as a function of Operation (addition, subtraction), Calculation Type (carry, no-carry, zero) and Required Force (null, low, high) factors. Continuous lines represent the high level of the Required Force factor, while dashed and dotted lines represent respectively the low and null level of the same variable.

As can be seen in figure 3, the constant error seems greater after addition than after subtraction for no-carry and zero types of calculation in both low and high Required Force conditions, which is a clear evidence of OM effect. On the other hand, in the null Required Force conditions (no matter which type of calculation), and in all the conditions of the carry type of calculation, the constant error shows trends incompatibility with the OM effect. To test our hypotheses, the constant error was submitted to a three-way repeated-measures ANOVA, including the factor Required Force (null, low, high) as the between-participants variable, and factors Calculation Type (carry, no-carry, zero) and operation (addition, subtraction) as the within-participants variables. The factor Calculation Type reached significance with $F(2, 204)=8.69$, $p<.001$, $\eta_p^2 =.07$ (carry=44.89, no-carry=42.55, zero=44.53). More interestingly, two interactions reached significance, i.e. the interaction between Required Force and Operation, with $F(2, 102)=5.95$, $p<.01$, $\eta_p^2 = .10$ (figure 4), and the interaction between Calculation Type and Operation, with $F(2, 204)=16.42$, $p<.001$, $\eta_p^2 =.13$ (figure 3).

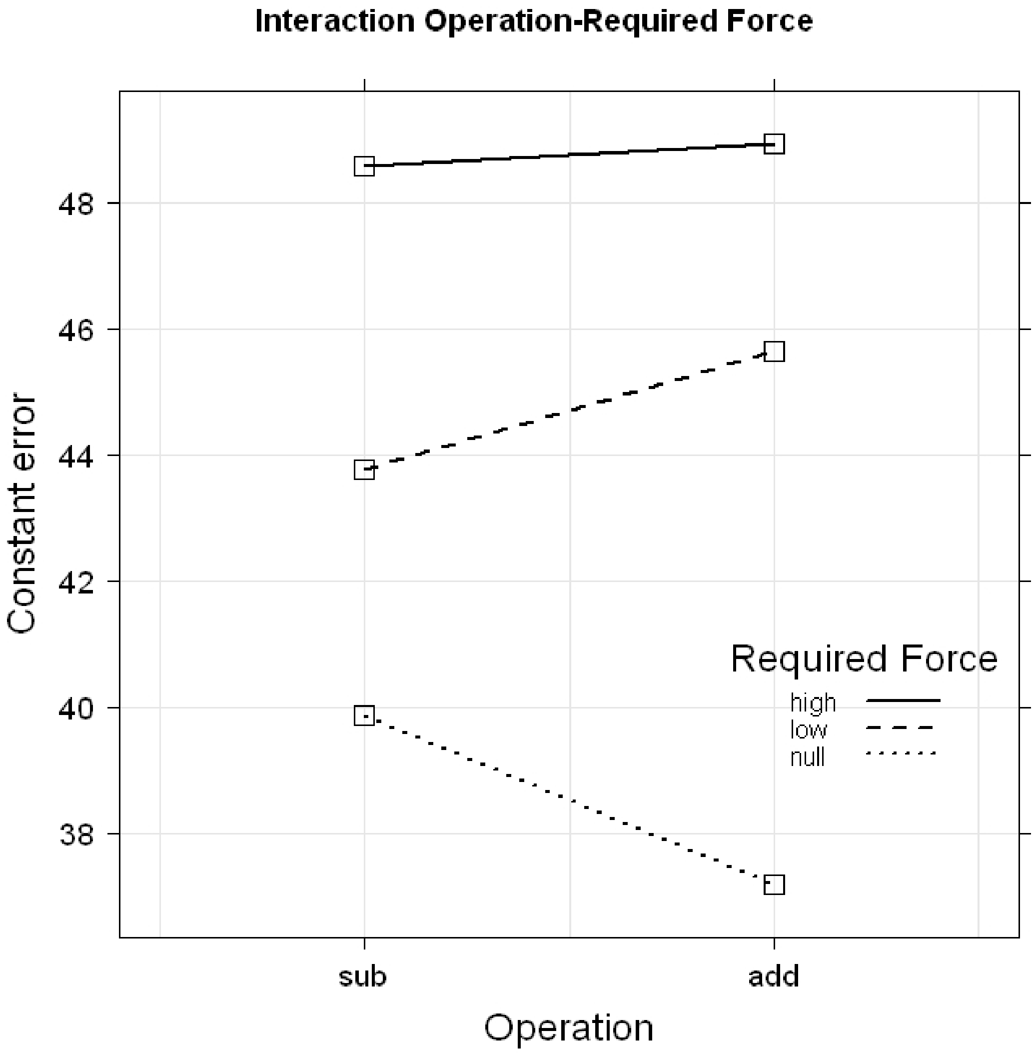

Figure 4: Interaction between Operation (addition, subtraction) and Required Force (null, low, high) factors. Continuous lines represent the high level of the Required Force factor, while dashed and dotted lines represent respectively the low and null level of the same variable.

As can be seen in figure 4, the interaction between Required Force and Operation seems to indicate that the grater overestimation in additions compared with subtractions depends on the force required, in that it reverses when no force is required. Moreover, the interaction between Calculation Type and Operation (see figure 5) seems due to the different trends of the carry type of calculation compared with the no-carry and zero types of calculation.

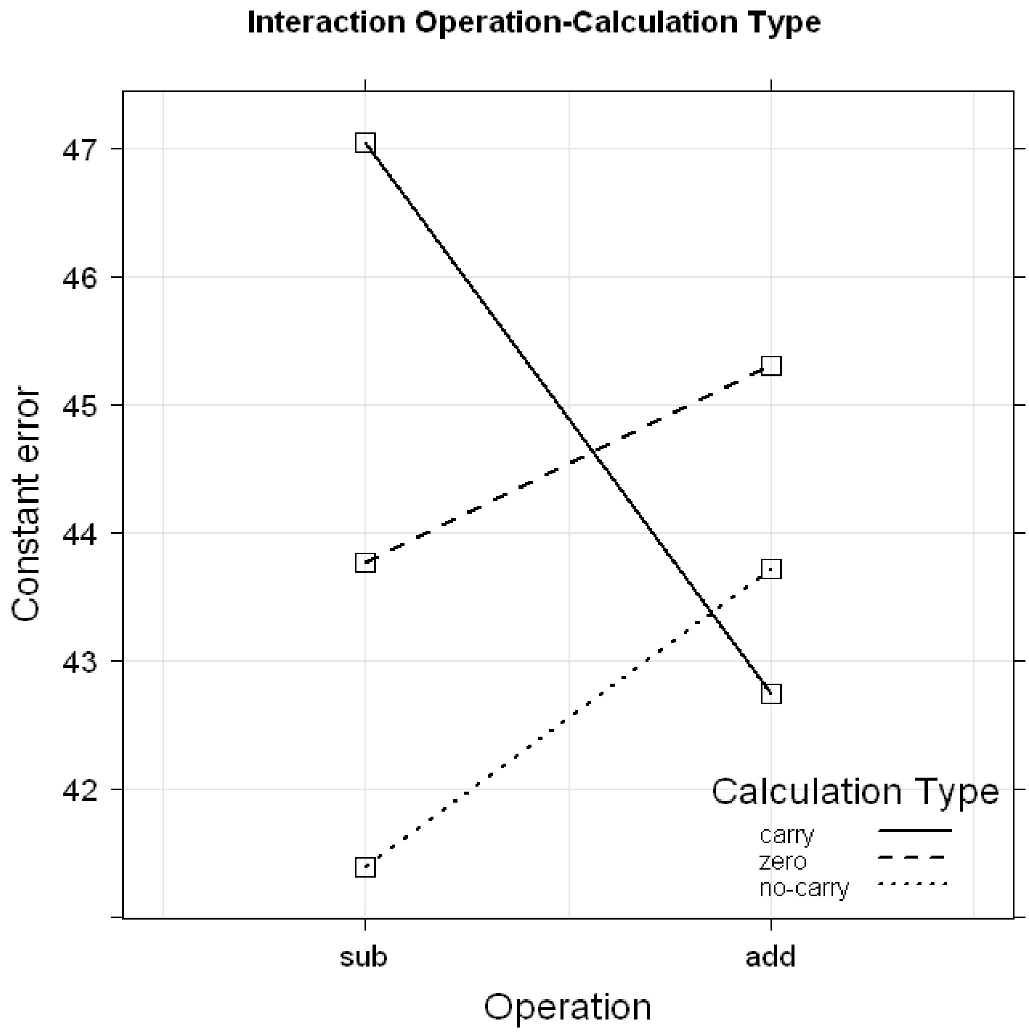

Figure 5: Interaction between Operation (addition, subtraction) and Calculation Type (carry, no-carry, zero) factors. Continuous lines represent the carry level of the Calculation Type factor, while dashed and dotted lines represent respectively the zero and no-carry level of the same variable.

Thus, considering that either the force required or the type of calculation seems to modulate the OM effect, two new ANOVAs were carried out, one for the carry type of calculation and one for the no-carry and zero types of calculation, to investigate the role of the force required in the different types of calculation. In the former ANOVA, with the between-participants factor Required Force (null, low, high) and the within-participants factor Operation (addition, subtraction), only the factor Operation reached significance with $F(1, 102)=16.21$, $p<.001$, $\eta_p^2=.13$ (addition= 42.74, subtraction= 47.05). The interaction failed to reach significance. In the ANOVA, for the other two Calculation Type conditions, with one between-participants factor Required Force (null, low, high) and two within-participants factors, i.e Calculation Type (no-carry vs. zero) and Operation (addition vs. subtraction), the significant sources of variance were Calculation Type, with $F(1, 102)=10.11$, $p<.01$, $\eta_p^2 = .10$ (zero=44.53 vs. no-carry=42.55) and Operation, with $F(1, 102)=11.37$, $p<.01$, $\eta_p^2 = 0.09$ (addition=44.51, subtraction=42.57). The interaction between Required Force and Operation was significant, too, with $F(2, 102)=5.24$, $p<.01$, $\eta_p^2=.09$, whereas the interaction between Calculation Type and Operation failed to reach significance ($F(2, 102)<1$).

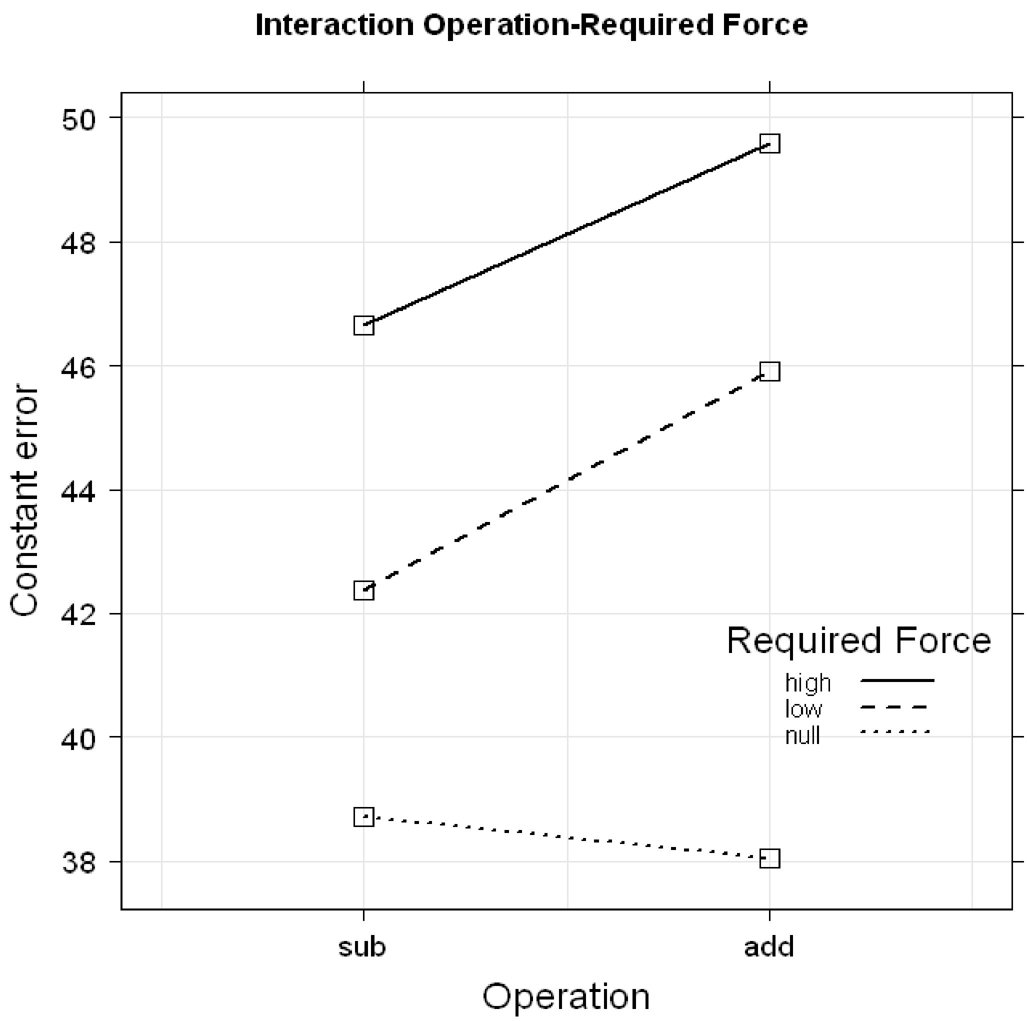

Figure 6: Interaction between Operation (addition, subtraction) and Required Force (null, low, high) factors in the ANOVA without the carry level of the Calculation Type factor. Continuous lines represent the high level of the Required Force factor, while dashed and dotted lines represent

As can be seen in figure 6, the trends show that the expected overproduction in addition compared to subtraction occurs in the high and low Required Force conditions only, and disappears when the force required is null. Indeed, planned comparisons confirmed that the difference between addition and subtraction is significant in the high and low Required Force conditions together ($F(1, 102)=21.22$, $p<.001$, $\eta_p^2 = 0.17$), whereas no significant difference was found between addition and subtraction in the null Required Force condition $F(1, 102)<1$. Moreover, the last planned comparison, carried out to check for parallelism of the trends of the high and low Required Force conditions, confirmed that the difference between addition and subtraction in the high Required Force condition does not differ from the difference between addition and subtraction in the low Required Force condition ($F(1, 102)<1$).

Discussion and Conclusions

The current study investigated whether the operational bias can be accounted for by a shared magnitude mechanism. Participants were instructed to solve addition and subtraction problems and generate a dot cloud as a response to the rotation of a knob. The crucial manipulation consisted of a between-subject variation of the force that is required to rotate the knob (null, low and high Required Force). Based on the idea that numerical magnitude and sensorimotor magnitudes are linked by a common system (Lindemann, Abolafia, Girardi, & Bekkering, 2007), a positive correlation was predicted between the operational bias and Required Force. Importantly, in line with our prediction, the operational bias in the low and high force group was larger than the operational bias in the null force group. This finding supports the idea of a functional role of shared magnitude codes for solving mental addition and subtraction.

The modulation of the operational bias by required motor force shows how the operational bias effect is influenced by the presence of a non-spatial sensorimotor magnitude, which it is difficult to explain by a mere spatial shifts of attention on the mental number line. Importantly, an operational bias was observed only when numerical magnitude was positively correlated with motor force. In the no force condition, in which numerical magnitude was not correlated with motor force, the operational bias was absent. Our results extend earlier evidence for the involvement of shared magnitude codes in number processing (Andres, Davare, Pesenti, Olivier, & Seron, 2004; Lindemann, Abolafia, Girardi, & Bekkering, 2007; Vierck, & Kiesel, 2010) to mental calculations. The fact that the operational bias is sensitive to demands on force processing supports the idea that a shared magnitude system is functionally involved in mental arithmetic.

It might be objected that since the knob had to be rotated clockwise to increase the number of dots, activation of the mental number line is driving the operational bias. Importantly, the operational bias was absent in the null force condition and only significant in low and high force conditions. If the rotation direction was responsible for our findings, the operational bias should also appear in the null force condition. However, our findings do not argue against the mental number line account of the operational bias, but suggest that the explanation may be more complex, involving a general representation of magnitudes. A study from Pinhas and Fischer (2008) has provided more direct evidence for a spatial mechanism underlying the operational bias. The authors found that participants locate the position of a number slightly leftwards after subtractions and slightly rightward after additions, compared to its actual position. An interesting question for future research is how far attentional shifts in space and processes within a shared magnitude system are related. According to the idea of shared representations for space, number, and sensorimotor magnitudes, it might be envisioned that both couplings of mental arithmetic with space and force are driven by a common system (Walsh, 2003).

Aside from the modulation of the operational bias by required motor force, the results from the present experiment are congruent with results from previous studies. First, the good fits of the estimations suggest that participants had no difficulties with our tasks; moreover, the subjects in both tasks responded quickly, and their response times do not increase with the number size, which reveals that they did not use a counting strategy but, instead, based their judgments on approximate numerosity estimation (Akin and Chase, 1978; Mandler and Shebo, 1982; Trick and Pylyshyn, 1993).

Second, in the numerosity judgment task, it was found that dots were systematically overproduced compared to the target, and so the number of dots in the visually presented set were systematically underestimated. The overproduction tendency was verified by a linear regression on log-transformed values with results in line with previous studies (Izard & Dehaene, 2008).

Third, in the mental calculation task, it was found that precision of judgments decrease with number size. Other studies on non-verbal approximate number processing have previously reported that variability estimates increase proportionally to number magnitude, a property called 'scalar variability' (Whalen, Gallistel, & Gelman, 1999; Cordes, Gelman, Gallistel, & Whalen, 2001; Izard & Dehaene, 2008). The present study confirms the notion that accuracy of number representations decreases with number size and provide, thus, new support for the adherence of internal number representations to the Weber’s law.

Fourth, in the mental calculation task the analysis of the first sub-design, shows that participants produced systematically more dots when indicating the outcome of an addition problem than when indicating the result of a subtraction problem, thus, providing empirical evidence of operational bias effect. This result replicates the finding of a response bias after cross-notational arithmetic as previously demonstrated (Phinhas and Fisher, 2008; Lindemann and Tira, 2001).

The analysis on performance in the mental calculation task also showed that the operational bias effect was not present if the operations required a carry or borrow operation; the difference between addition and subtraction operations was not only reduced for carry operations, but even reversed. Carry operation involves a decomposition of the place-value system and results in an increased load of phonological working memory resources (DeStefano & LeFevre, 2004; Deschuyteneer et al., 2005; Kalaman & Lefevre, 2007; Imbo et al., 2007). Therefore, one might assume that the processing of carry operations, which is strongly based on verbal processing strategies, engages fewer non-verbal analog representations of numerical magnitude information. Furthermore, this could explain the lack of effect of the variable Required Force on Carry operations. An alternative explanation for the absence of the operational bias effect for carry operations could be that participants ignored or approximated the unit values of the proposed two digit numbers. Considering carry operations, in the case of an addition, this heuristic would result in an underestimation of the outcome, and in the case of a subtraction in an overestimation. However, this approximation would leave the difference between addition and subtraction in no-carry and zero operations virtually unaffected on average. Considering the mean result of the operations presented in this work against the mean values resulting by the application of this heuristic (Figure 7), it is possible to see how this could explain the inverse OM effect in carry operations.

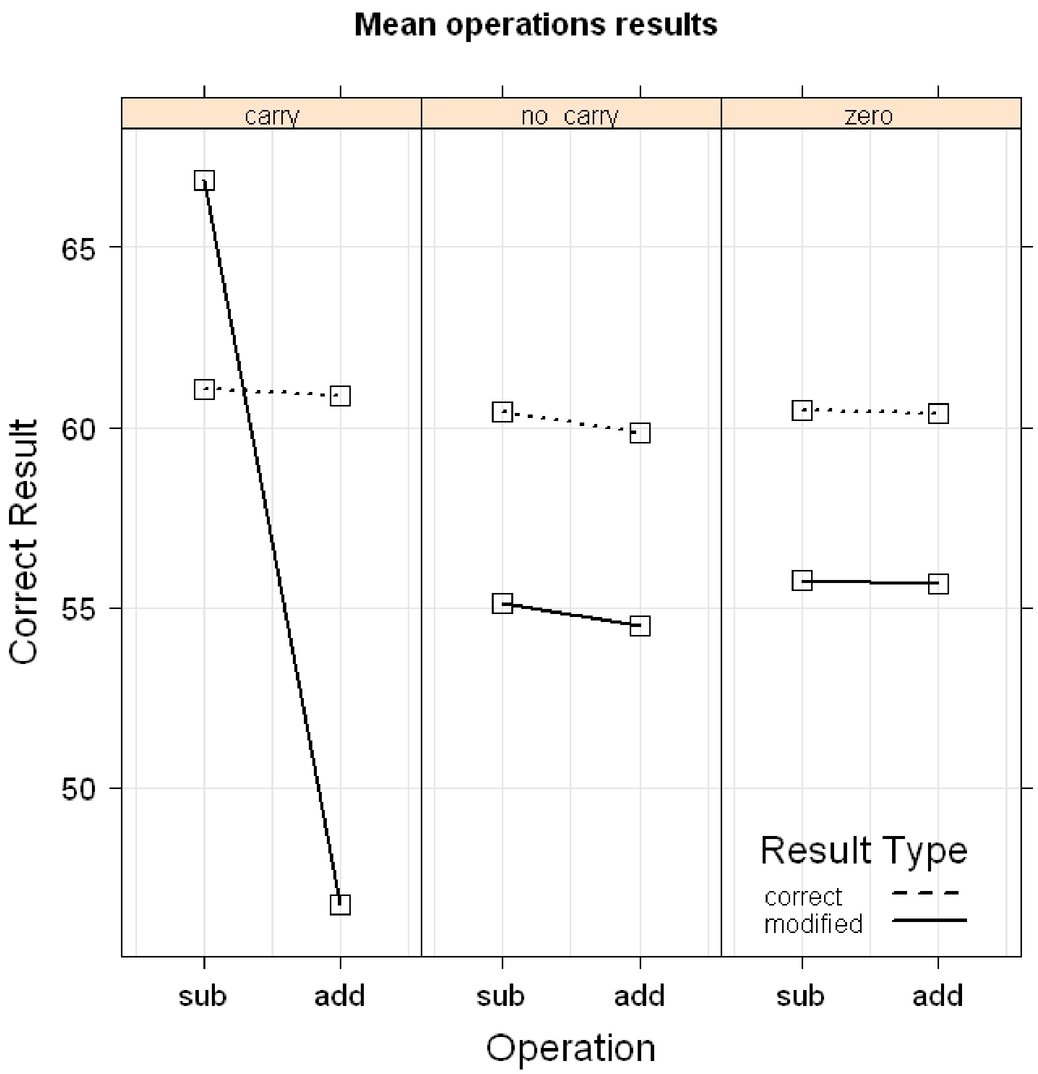

Figure 7: Plot of the operations' results in the calculation task divided by Operation (addition, subtraction) and Calculation Type (carry, no-carry, zero). Dashed lines represent the correct outcome of the operations, while continuous lines represent the modified operations' results, computed ignoring the units value of the two operands.

Number magnitude has been demonstrated to be connected to several domains, as, space (e.g., Dehaene et al., 1993) and time (e.g., Xuan et al., 2007) for example. Walsh (2003) proposed a system located in the inferior parietal lobe, in which magnitudes of different domains are represented together and are used as the basis for action. Within this framework, the operational bias effect has been interpreted as support for an analog representation of symbolic and non-symbolic numerical magnitude information (McCrink et al., 2007; Pinhas & Fischer, 2008). Our findings extend the existing knowledge on magnitude interferences to mental arithmetic, supporting the idea of an involvement of the AMS in mental arithmetic.

Study 3: Individual differences in the representation of numbers

As seen in the introductory sections, one of the most influential representations of numbers takes the form of a mental number line. Different behavioral measures are assumed to tap into this representation and to reveal different aspects of it. One of these measures, by showing an association between sides of space and magnitude of numbers, is thought to reflect the spatial orientation of the MNL, that is, the SNARC effect. The distance effect is used to characterize the precision of the MNL by assessing the degree of overlap between the representations of two different numerosities. Behaviorally, as seen in the introduction section, the distance effect corresponds to slower reaction times when comparing two numbers separated by a smaller numerical distance. The more precise the MNL, the less participants are impacted by the numerical distance between the numbers to be compared, resulting in a smaller distance effect. the distance effect was investigated in the context of symbolic comparison tasks

Interesting example of within magnitude interferences commonly explained with the NML is the *size-congruency effect*. Henik and Tzelgov (1982) showed that when Arabic numerals physical size is systematically manipulated, judgments of which numeral is larger (in size or number) showed congruency effects between the attended and unattended dimensions.

Studies that have examined the SNARC effect, as well as other well-established effects tap into the MNL representation, focused more on the cognitive mechanisms that underlie these effects, and less on individual differences in these measures of numerical representations. Because, as seen before in this work, the MNL hypothesis and the mechanisms of the SNARC effect are still debated, a better understanding of how individual differences in the SNARC relate to individual differences in numerical and arithmetical processes may shed light on the underlying mechanisms of Number representation. Viarouge, Hubbard and McCandliss (2014) verified the stability of the individual differences of SNARC effect on diverse sessions, that is, subject tested in two sessions sowed similar effects. If the SNARC does indeed represent a stable feature of individuals cognition, then it would be important to know how it relates to other cognitive processes, such as numerical comparison and mental calculation.

The aim of studying the relationship of the SNARC effect to other numerical (distance effect and size-congruency effect) and arithmetical (operational bias effect) measures, was to investigate the cognitive mechanisms underlying those measures with an individual differences approach. That is, assuming that the SNARC, the distance effect, the size congruency effect and operational bias are all behavioral indicators of a unified mental number line, these measures should share common variance from a participant to an other. The correlation between all those effects suppose to be signature for the MNL has not yet been directly tested. To date, only two studies concerning the correlation between the SNARC and the distance effect has been performed (Schneider, Grabner & Paetsch, 2009; Viarouge, Hubbard & McCandliss, 2014). Schneider, Grabner and Paetsch (2009) conducted a study on children that yielded contradictory results concerning the correlation between the SNARC and the distance effect. Viarouge, Hubbard & McCandliss (2014) studying the relation between SNARC and the distance effect in adults, reported a significant correlation between the two effects. The aim of this study was to broaden the existing knowledge about the connection among those effects seeking for a underlying behavioral sign of a common processing.

Method

Participants. Thirty-six undergraduate students from the Padova University (14 males) participated in the experiment. Participants were all right-handed, native Italian speakers, all reported having normal or corrected-to-normal vision. The average age of participants was 24.19 years (SD = 5.71). Participants received no compensation for their participation and were naıve to the study hypotheses.

Tasks. Three tasks were used: a parity task, odd/even judgment on visually presented Arabic digits; a symbolic comparison task, comparison of a visually presented Arabic number to a reference number; and a mental calculation task with a numerosity production response method, solve an arithmetic problem presented in Arabic format via the production of a non symbolic numerosity. In all three tasks, participants were instructed to give their response as quickly and as accurately as possible.

Parity task. Stimuli were Arabic digits between 1 and 9, excluding 5, presented in Microsoft Sans Serif font 26 pt. All stimuli were presented in the middle of the screen, in withe on a black background. On each trial, participants were asked to indicate the parity of the presented number by pressing either the most leftward or the most rightward button of the response box using their left and right index fingers. For each trial, a fixation point (a cross of 26 pt font size) appeared in the center of the screen for 1 s, followed by the target digit, which disappeared as soon as participants responded or after 1.5 s. There was an inter-trial interval (ITI) of 1 s between each experimental trial. The Parity task section was divided into four blocks between which the participants were allowed to take a short break. The position of the fingers on the buttons remained constant throughout blocks and across participants, that is, right index finger on the right button, left index finger on the left button. The assignment between the parity and the response buttons was switched between blocks, and the order of the blocks was counterbalanced across the participants. Blocks 1 and 3 were designed as training blocks to help reinforce the response button mapping; here each of the 8 digits were presented twice in a random order. In the experimental blocks (2, 4), each of the 8 digits were presented 9 times, resulting in a total of 72 trials per block, randomized within each block and with the same parity-button assignment of the respective training block (1-2, 3-4).

Comparison task. The design of the comparison task was similar to the design of the parity task, with some exception: each of the 8 digits were presented 10 times per experimental block, half of the time in 26 pt font and the other half in 36 pt font, comprising a total of 80 trials per experimental block. Moreover, participants were asked to decide as quickly and accurately as possible whether the presented digit was more or less than 5 by pressing the left or right response button with the corresponding index finger. The order of the assignment between the responses and the buttons was counterbalanced across participants and across the two groups defined by the parity task button-mapping orders (Table 2).

Table 2. Experimental design. Each session began with a parity task, followed by a comparison task and an approximate arithmetic task.

Sequence 1	Parity-Sequence 1	Comparison-Sequence 1	Calculation
Sequence 2	Parity-Sequence 1	Comparison-Sequence2	Calculation
Sequence 3	Parity-Sequence 2	Comparison-Sequence1	Calculation
Sequence 4	Parity-Sequence2	Comparison-Sequence2	Calculation

Mental calculation task. The design of the mental calculation task was similar to the design presented in Study 2, with some exception: only no-carry operations and low required force have been used. The mental calculation task section was divided into two blocks between which the participants were allowed to take a short break. Block 1 was designed as training block to help reinforce the numerosity production response method; here 20 random two-digits numbers were presented with the request to produce such a numerosity via the numerosity production response method. In Block 2, for each trial an addition or subtraction problem was shown, the participant was instructed to indicate the result of the presented problem as quickly and accurately as possible. A list of 24 addition and 24 subtraction operations has been compiled for a total of 48 trials (see Table 3). Each problem comprised two operands. No decade numbers or symmetric numbers (e.g., 22, 33) occurred as operand or result. Operands and result of one problem never had the identical decade or unit digits. For additions and subtractions, half of the first and half of the second operands were odd numbers. The outcomes of addition and subtraction operations were matched.

Table 3: Addition and Subtraction operations used in the mental calculation task. The results were not presented during the experiment.

Addition operations	Subtraction operation	Results
13 + 21	57 – 23	34
21 +14	69 – 34	35
12 + 24	98 – 62	36
14 + 23	58 – 21	37
12 + 41	74 – 21	53
24 + 32	68 – 12	56
41 + 17	89 – 31	58
53 + 14	79 – 12	67
51 + 32	97 – 14	83
63 + 21	97 – 13	84
12 + 73	98 – 13	85
52 + 34	98 – 12	86

Material and Procedure. Participants took part in a forty minutes experimental session. The three different tasks were programmed using E-prime software (Parity and Comparison tasks)(Schneider, Eschman, & Zuccolotto, 2012) on a Dell 32-bit personal computer, equipped with a 26’’ screen ;and with the Expyriment Python's package (Krause, Lindemann, 2013) on a Acer 3.33 GHz 64-bit personal computer, equipped with a 26’’ screen. The two computers were in the same room and the testing conditions (room and experimental set up) were kept constant across participants. A two buttons response box was used for response collection in both the parity and comparison tasks, while, in the approximate arithmetic task the same rotating knob keyboard used of Study 2 were in use.

Results

Standard regression analysis assumes independence between different observations; an assumption that is usually violated by data from within-subjects designs. This problem can be circumvented regressing a dependent variable on an independent variable individually for each participant and then comparing the extracted value for slopes between conditions or against a population value via standard significance tests. This procedure, commonly known as regression coefficient analysis (RCA; Lorch & Myers, 1990, Method 3), only assumes a linear relationship between dependent and independent variables for each individual participant and can be used for both, continuous and dichotomous predictors (Ahn, Jung, & Kang, 2002; Lorch & Myers, 1990; Myers & Broyles, 2000). Each of the effects considered in this experiment was tested using the RCA technique implemented with R statistical package (R Core Team, 2013) as described by Pfister, Schwarz, Carson, and Jancyzk, (2013)

Analysis of the SNARC effect. Correct trials with reaction times between 150 and 1200 ms were included in the analysis, as previously used in SNARC studies (Schwarz & Muller, 2006). For each participant, more than 86% of the total number of trials respected this criterion. For each participant and each of the 8 numbers tested were calculated the difference in reaction times between the position of the response-button (dRT = mean RT right – mean RT left). For each participant was then computed the slope of the linear regression of the 8 numbers on the dRTs, representing the amplitude of the SNARC effect (Fias, Brysbaert, Geypens, &

D'Ydevalle, 1996; Lorch, & Myers, 1990). Showing an advantage for left-sided over right-sided responses for small numbers and an advantage for right-sided responses for large numbers, a negative slope indicates the presence of a SNARC effect. A t-test was then performed on the obtained regression slopes to test whether the slopes were significantly different from 0. The Cohen's d effect size was calculated. The t-tests performed on the regression slopes showed a significant SNARC effect across the 36 participants (mean slope = -0.25, $t(35) = -3.60$, $p < .001$, $d = -1.21$). This result replicated previous findings and confirmed the presence of a global SNARC effect in our group of participants.

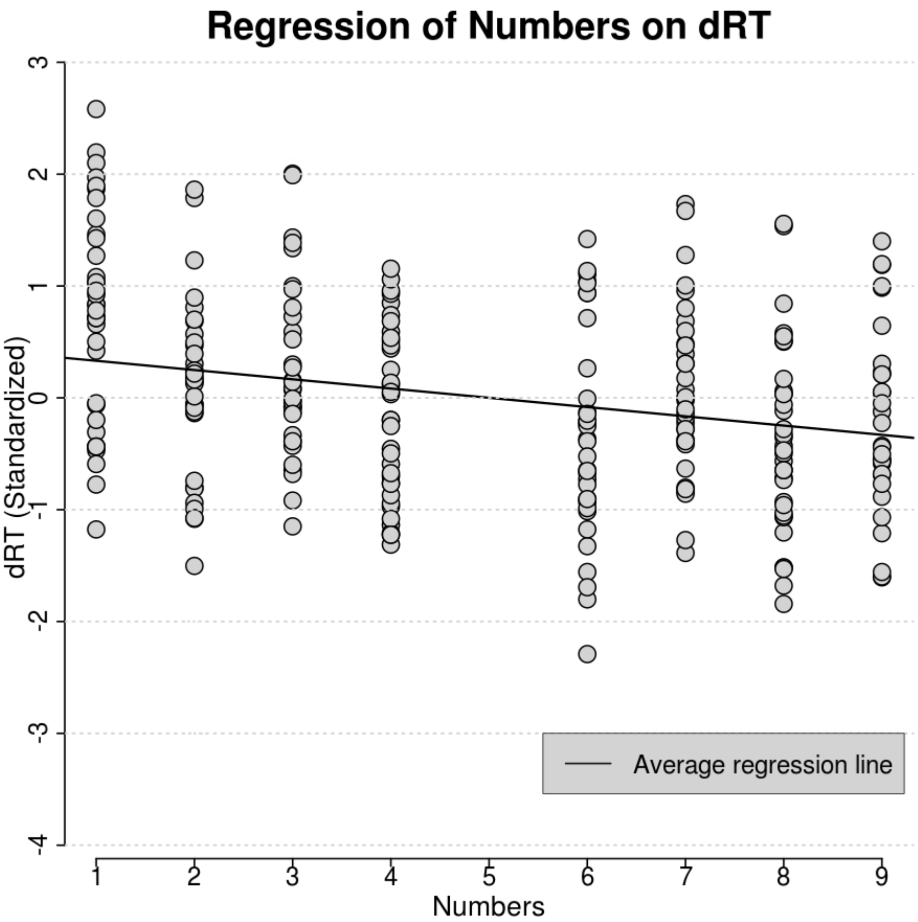

Figure 8: SNARC effects (RT right -RT left) as a function of the corresponding target number clearly show a negative slope.

Analysis of the distance effect and size-congruency effect. The data from the comparison task were analyzed to probe both the distance effect and the size-congruency effect. As with the parity task, only correct trials with reaction times between 150 and 1200 ms were included in the analysis. For each participant, more than 82% of the total number of trials respected this criterion. In order to calculate the amplitude of the distance effect, the trials were grouped based on the absolute value of the distance to the reference digit 5, and computed for each participant the average reaction time for the four distances (1, 2, 3 and 4). For each participant was then computed the slopes of the regression for the four distances, the two font sizes and their interaction, against the average reaction times. The slope of the distances represented the amplitude of the distance effect and the slope of the font size represented the amplitude of the size-congruency effect in each subject. In the first case, a negative slope shows a decrease in reaction times as the distance between the target and the reference digit increases, which is the distance effect. In the second case a positive slope shows a decrease in reaction times when the font size of the number matched the relation between the target and the presented number, which is the size-congruency effect.

All participants showed a decrease in their average reaction time with increasing numerical distance. A t-test performed on the 36 slopes showed a significant distance effect across our group of participants (mean slope = -0.36, $t(35) = -4.09$, $p < .0.001$, $d = -1.38$).

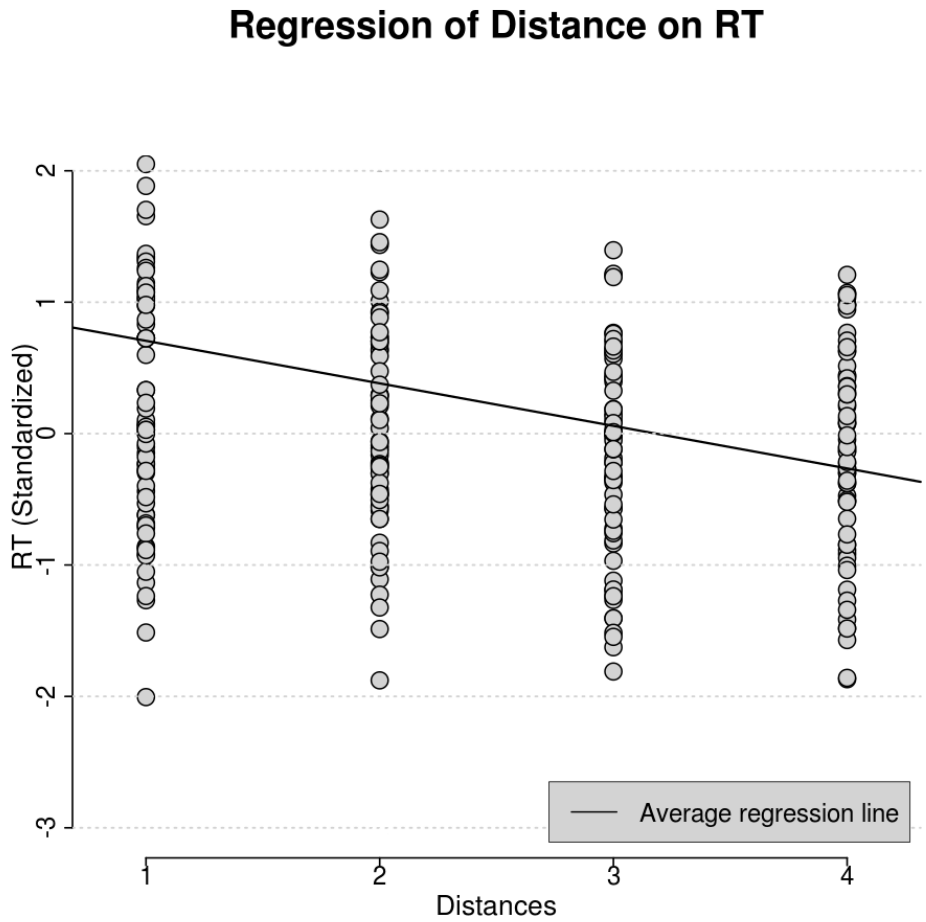

Figure 9: Reaction times as a function of the corresponding distance of the stimulus number from the target number clearly show a negative slope, that is the distance effect.

Moreover, all participants showed a decrease in their average reaction time when the font size of the presented number matched the relation with the target number. A t-test performed on the 36 slopes showed a significant size-congruency effect across our group of participants (mean slope = -.44, $t(35) = -3.82$, $p < .001$, $d = -1.29$).

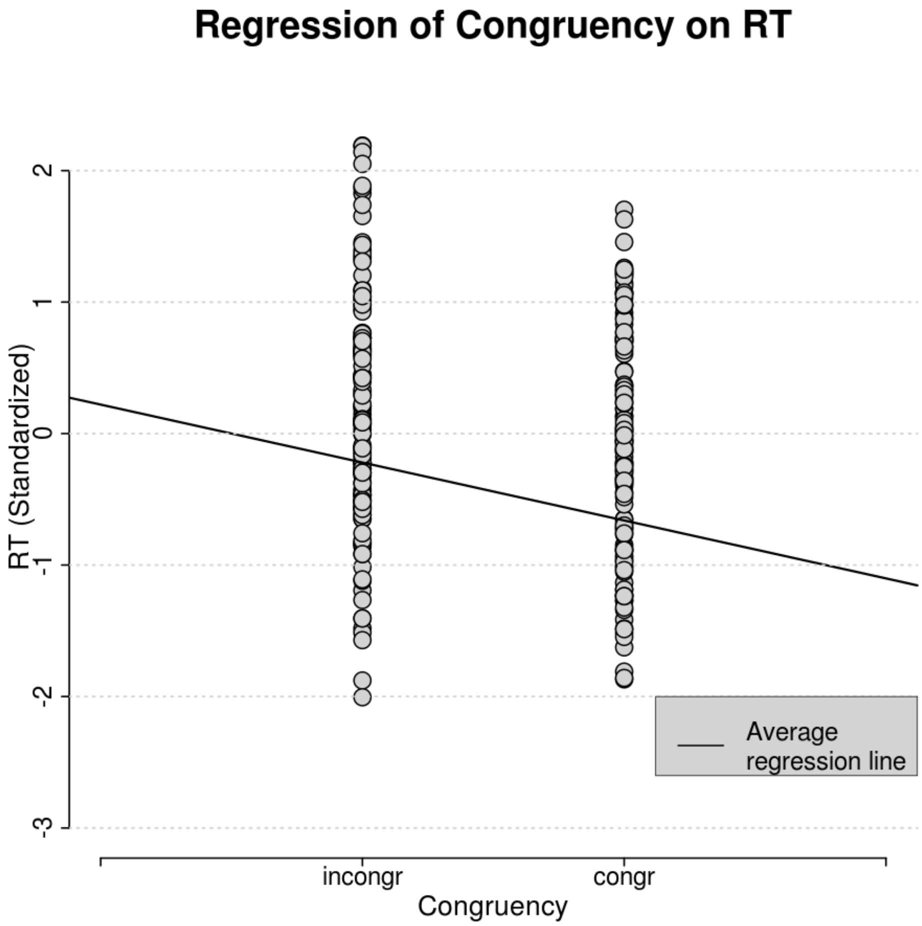

Figure 10: Reaction times as a function of the congruency conditions of the stimulus clearly show shorter RT for the congruent condition than for incongruent, that is the size congruency effect.

Analysis of the operational bias effect. Trials with estimations equal to zero or three time away from the subject average estimation were not included in the analysis. For each participant, more than 91% of the total number of trials respected this criterion. For each participant was calculated the constant judgment error, defined as the average error between correct outcomes and responses (Schutz & Roy, 1973). For each participant was then computed the slope of the linear regression of the operation on the constant error. Showing higher estimates for addition compared with subtractions, a positive slope indicates the presence of a operational bias effect. A t-test were then performed on the obtained regression slopes to test whether the slopes were significantly different from 0. The t-tests performed on the regression slopes showed a significant operational bias effect across the 36 participants (mean slope = -.20, $t(35) = -4.46$, $p < .0001$, $d = -1.50$). This result replicated previous findings and confirmed the presence of a global operational bias effect in our group of participants.

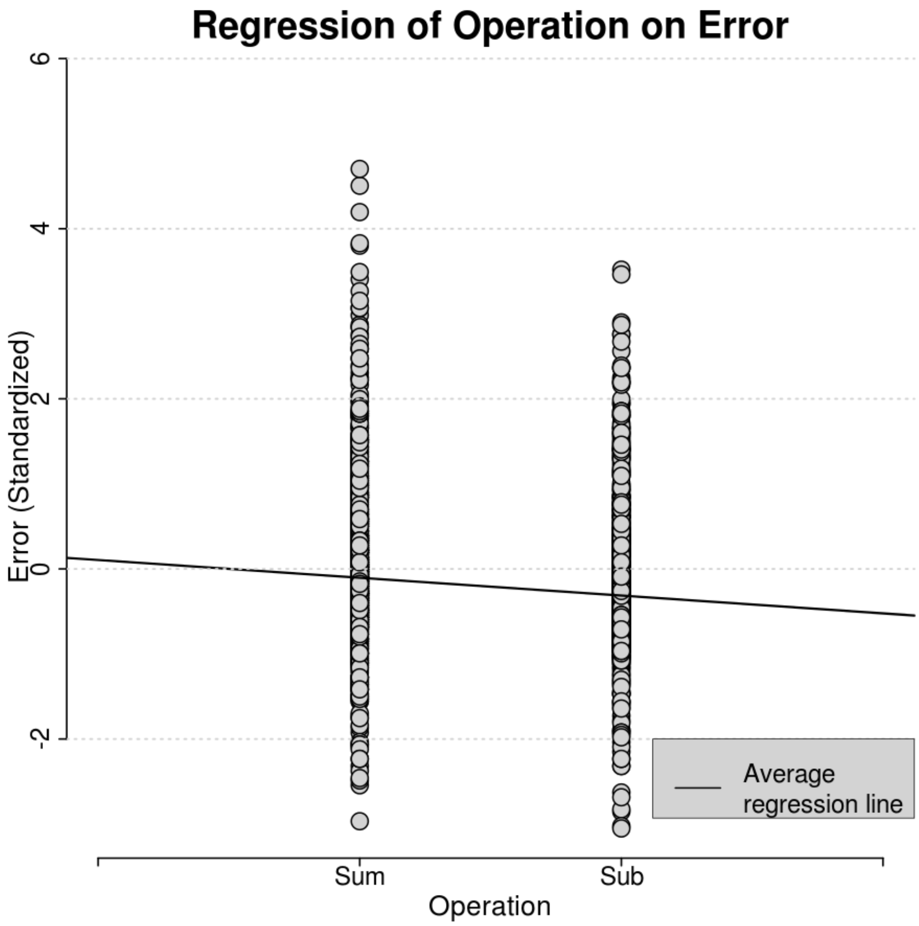

Figure 11: Error (estimation - target) as a function of the operation (+, -) of the stimulus arithmetical problem, clearly show higher overestimation for addition than for subtraction, that is the operational bias effect.

Links between the different tasks. The correlation between the SNARC effect, the distance effect, the size-congruency effect and the operational bias effect was tested using the Pearson product-moment. The standardized amplitude of each of the effects, calculated at a participant level, was regressed against the same measure of each of the other effects to explain the variance in the mean amplitude of each of the others correlation coefficient (Pearson, 1895).

As can be seen in Table 3, the only two significant correlations can be observed between the mean amplitude of the SNARC effect and the amplitude of the congruence effect (Pearson's r = -.43, p < .05), and between the mean amplitude of the distance effect and the amplitude of the congruence effect (Pearson's r = .66, p < .05).

Table 3. Pearson product-moment correlation coefficients and relative probability values.

	Distance	Congruence	Operational_Bias	SNARC
Distance		0.66 (p < .001)	-0.07 (p = .68)	-0.31 (p =.07)
Congruence			-0.10 (p = .55)	-0.43 (p <.05)
Operational_bias				-0.30 (p =.08)
SNARC				

Discussion

The present study intended to test if the SNARC effect, the distance effect, the size congruency effect and operational bias share common variance from a participant to an other, thus indicating their belonging to a stable, task independent, mental representation of numbers such as the mental number line. Basic step to study the relationship between different cognitive effects on a subject level is to find those effects in the set of participants.

During the parity task, participants were instructed to press one of two buttons if a presented number was even and the other button if odd. The results show that, on average, responses to small numbers (1,2,3, and 4) given with the left hand and to big numbers (6,7,8, and 9) with the right, were faster than responses to big numbers given with the left hand and to small numbers with the right. That is, on average participants showed a statistically significant SNARC effect.

During the comparison task, participants reported whenever a sown number was bigger or smaller than the target number 5. The data show how, on average, the bigger was the distance between the presented and the target number, the faster was the response. Thus, participants showed statistically significant distance effect. Moreover, during the magnitude comparison task the numbers to be compared were presented alternatively in two font size. The font size of the number was an irrelevant dimension for the completion of the task, nevertheless the result shows shorter reaction times when the font size of the stimulus number was congruent with the numerical value of the presented number in respect to the target, that is, when a number smaller than the target was presented in small fount and vice versa.

In the last of the three tasks, the mental calculation task, participants were asked to mentally calculate the result of a symbolically presented addition or subtraction, and to express that result controlling the numerosity of a dot pattern. The results show how bigger estimations followed additions and smaller estimations followed subtraction, even if the true outcomes between them were matched. This discrepancy in the estimation of the results between addition and subtraction is the operational bias.

Summarizing, the participants showed a SNARC effect, a distance effects, a size congruency effect, and an operational bias effect, replicating all together the result of previous studies.

The multiple regression analysis operated on the individual amplitude of each one of the considered effects reveals that the relative contributions of SNARC effect and distance effects represent significant, yet distinct, contributions in explaining variation in the size of the size-congruency effect on a participants level. Differently from what found by Viarouge, Hubbard and McCandliss (2014) no statistically significant correlation was found between the SNARC and the distance effect. Our results show that participants with a stronger size congruency effect have weaker SNARC effect and a stronger distance effect. Whereas, SNARC effect do not show statistically significant correlation with the distant effect.

More interesting the operational bias effect seem to share variance with none of the considered numerical effects. This may be due to the increase complexity involved in a calculation task compared to a number task.

Several studies using transfer paradigms (Proctor, Yamaguchi, Zhang, & Vu, 2009; Yamaguchi & Proctor, 2009) reveal the influence of task-defined mappings on processing of task-irrelevant stimulus attribute in sequential tasks. That is, when response selection is performed based on associations between specific stimuli and responses, after an associations are learned, these associations remain in memory and affect performance in subsequent tasks if those particular stimulus features occur and retrieve the learned associations (Bae, Choi, Cho, & Proctor, 2009).

To test the possible effect of mapping sequence from a task to the sequent one, the regression coefficients derived from the magnitude comparison task and from the mental arithmetic task should have submitted to separate two-way ANOVAs, including the factors parity task mapping sequence (1,2) and magnitude comparison task (1,2) as the between-participants variables. Nevertheless, the small sample size for each one of the resulting four, between participants, design cells (9) would be inadequate to perform such tests.

Thus, it is only possible to speculate about an influence of the mapping sequence on the result of both the magnitude comparison and mental arithmetic tasks. However, for completeness in the discussion of the results it will be shown how the correlation data seems to draw a picture that might have failed to get statistical support due to the possible influence of the mapping sequence. From this point of view the operational bias effect might correlate to the SNARC effect while the SNARC effect, the distance effect and the congruence effect might all correlate. It is possible that those results failed to get statistically significant due to the effects of the mapping seguence

This result would support the idea of a shared representation of numbers in number tasks; more over the correlation of the operational bias with the SNARC effect could support the idea of a shared representational mechanism between them. Nevertheless, the negative value of this correlation speaks against the idea of the NML as common basis for the two phenomenons.

To clarify the relationship between the tested effects with better confidence a new experiment based on the experience of this study would be in order.

Chapter 3

General Conclusion and Discussion

The present work showed a wide corpus of scientific literature about adult humans sharing with non-human animals a non-verbal system for representing discrete and continuous quantity. These quantities have the properties of continuous magnitudes, that is, errors in estimations about magnitudes are not equiprobable but tend to distribute as a Gaussian curve around the estimation. The brains of tested human and non-human animals perform arithmetic operations with mental magnitudes; they add, subtract, multiply, divide and order them. The processes that map numerosities (discrete quantities) and magnitudes (continuous quantities) into mental magnitudes. Moreover, the operations that the brain performs on those mental magnitudes, lead to approximate, but still valid results used to effectively drive behavior. From those considerations, it is reasonable to think of a neural substrate for this system, evolved far back in time to drive action in a complex environment.

Despite these notions, a lot abut the fine nature of the mental representation of number is still unclear; two general accounts was described for clarifying and building hypotheses about the mental representation of number and its overlapping with the representations of space: the mental number line hypothesis and the common magnitude system hypothesis. The first one assume the analog numerical magnitudes to be represented via a positional coding along a spatial continuum in which numerical magnitudes are mapped onto mental space from left to right in ascending order. Instead, the second representational system, would represent numerical magnitudes in a format, analogous to continuous magnitudes, shared with other magnitudes, like time, space, and sensorimotor magnitudes. This second position assume that, even if the findings about the within magnitude interferences tend to emphasize the relations between the dimensions of space and number. The fact that there are so many studies that report a relationship between those dimensions, and not others, can't by itself led to the impression that there is a phylogenetically privileged relationship among the dimensions of space, time, and number (Dehaene, Izard, Spelke, & Pica, 2008; Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999; Srinivasan & Carey, 2010; Walsh, 2003). Indeed, many studies shows evidence for interaction among quantitative dimensions beyond space and number that can hardly be explained by the MNL hypothesis.

The experimental works here presented build up from this knowledge to clarify and refine those models of number representation. The first study was composed by two experiments and its aim was two folded; first to validate a new response method to be used in numerosity and arithmetical tasks (numerosity production response method). This was done comparing the results derived with this new method to the results from a classical method of response (categorical scaling).

The second aim was to compare the influence of primacy and recency between the two operands on the estimated result of an averaging operation. In both experiments, the reaction times of the participants were short and did not increase with numerosity, that is, responses were not produced using counting strategies, that is important for the validity of both the results and the new response method. In both of the experiments, the results of the analysis on the estimated averaging values seemed to indicate that the weights of the two dot patterns did not differ significantly. As a consequence, the scale values were affected only by the numerosity of the stimulus and not by its position (1st or 2nd dot pattern). This result demonstrates that neither the effect of primacy nor the effect of recency influence the evaluation of the average numerosity, despite the sequential temporal order of the presentation of the stimuli (Busemeyer, 1991). In other words, this means that the participants give the same importance to the two quantities of each trial during averaging operations. This result legitimate the use of sequential presentation of stimuli in non-symbolically presented arithmetical tasks, reducing the fear of primacy and recency effects. Since the participants were instructed to perform an averaging operation one might object on the generalizability of this result to other arithmetic operations. The averaging operation was chose because at the moment was the only one enabling the evaluation of the weight between the operands. Future development of IIT framework will hopefully provide with tools able to evaluate the operands' weights even in additions and subtractions.

Since an averaging operation was performed, if the response measure was on a linear scale, the factorial plot should exhibit parallelism. According to the ANOVA on the full factorial design, both the rating data (Figure 1) and the numerosity production data (Figure 2) show clear parallelism, thus validating the numerosity production, as a response measure on a linear scale, a prerequisite for a method to study stimulus interaction, and for the analysis of non-linear integration rules (Anderson, 1982). Moreover, as a consequence of those results the IIT framework was established for the first time as a viable instrument in testing mental arithmetic with discrete quantities.

Aim of study 2 was to investigated whether a shared magnitude mechanism can account for the operational bias. This experimental question started from the idea of a common system linking numerical magnitude and sensorimotor magnitudes. In a mental arithmetic task using a numerosity production method of response, the force required to rotate the knob was manipulated in a three between participants levels. A positive correlation between the operational bias and the factor Required Force was expected. In line with our prediction and with the results of Lindeman et al. (2007), the operational bias in the low and high force group was larger than the operational bias in the null force group. This finding supports the idea of a functional role of shared magnitude codes for solving mental addition and subtraction.

The modulation of the operational bias by the required motor force shows how the mental representations underling the operational bias effect are, at some stage of the processing, influenced by the presence of a non-spatial sensorimotor magnitude. Importantly, an operational bias was observed only when numerical magnitude was positively correlated with motor force. In the no force condition, in which numerical magnitude was not correlated with motor force, the operational bias was absent. Our results extend earlier evidence for the involvement of shared magnitude codes in number processing (Andres, Davare, Pesenti, Olivier, & Seron, 2004; Lindemann, Abolafia, Girardi, & Bekkering, 2007; Vierck, & Kiesel, 2010) That in, the sensitivity of the operational bias to demands in force processing supports the involvement of a shared magnitude system is functionally in mental arithmetic. Aside from the modulation of the operational bias by required motor force, the results from study 2 are in general congruent with results from previous studies.

Study 3 intended to test with an inter-individual differences approach a few important effects in the field of mathematical cognition. The SNARC effect, the distance effect, the size congruency effect and operational bias were tested together to investigate the possibility of them sharing common variance from a participant to an other, thus suggesting their belonging to a stable, task independent, mental representation of numbers. All participants took part in three different tasks, a parity judgment task where they had to discriminate between even and odd numbers, a magnitude comparison task where they were instructed to discriminate between number smaller or bigger than 5, and a mental arithmetic task, in which they had to judge the result of symbolically presented additions and subtractions. By means of these tasks, the presence of all the effects object of this study was verified in the participants set. Moreover, linear regression have been used to calculate the regression coefficient of each subject for each effect and test the correlation between the regression coefficients of all of them. The result of this study, even if not conclusive, point in the direction of a common representational mechanism underling the tested numerical effects (SNARC, size congruency, distance). On the over hand the operational bias seem to have a negative correlation only with the SNARC effect. This might suggest a connection between the two, but for the negative polarity of this correlation, the derivation of those effects from a stable mapping of numbers on mental space might seem improbable. Due to shortcomings in a piece of design in Study 3 a new study to clarify the relationship between the tested effects with an higher level of certainty is in order. Nevertheless Study 3 demonstrated the existence of an inter-individual correlation between the SNARC effects and the size congruence effect, and between the distance effect and the size congruence effect.

Altogether the present research work, by showing firstly non-spatial within magnitude interferences between number and a sensorimotor magnitude, and secondly a negative correlation between the SNARC effect and the operational bias, supports the idea of a more more complex account for the number representation than the NML hypothesis. Walsh (2003) proposed a system located in the inferior parietal lobe, in which magnitudes of different domains are represented together and are used as the basis for action. Within this framework, the operational bias effect has been interpreted as support for an analog representation of symbolic and non-symbolic numerical magnitude information (McCrink et al., 2007; Pinhas & Fischer, 2008). Our findings extend the existing knowledge on magnitude interferences to mental arithmetic, supporting the idea of an involvement of the AMS in mental arithmetic. Those findings do not argue against the mental number line account of the operational bias, but suggest that the explanation may be more complex, involving a general representation of magnitudes of which the mental number line could be a part. According to the idea of shared representations for space, number, and sensorimotor magnitudes, it might be hypothesized that both couplings of mental arithmetic with space and force are driven by a common system (Walsh, 2003). Future works will have to investigate how far attentional shifts in space and processes within a shared magnitude system are related.

References

Ahn, C., Jung, S. H., & Kang, S. H. (2002). Modified regression coefficient analysis for repeated binary measurements. *Journal of Applied Statistics* , 29 (5), 703-710. doi: 10.1080/02664760120098766

Andres, M., Davare, M., Pesenti, M., Olivier, E., & Seron, X. (2004). Number magnitude and grip aperture interaction. *Neuroreport*, 15(18), 2773-2777.

Akin, O., & Chase, W. (1978). Quantification of three-dimensional structures. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 397-410.

Allik, J. (1989). Is a unified psychophysical law realistic. *Behavioral and Brain Sciences*, 12(2), 267-268.

Allik, J., & Tuulmets, T. (1991). Occupancy model of perceived numerosity. *Perception & Psychophysics*, *49*(4), 303-314.

Anderson, N. H. (1962). Application of an additive model to impression formation. *Science, New Series*, *138*(3542), 817-818.

Anderson, N. H. (1965). Averaging versus adding as a stimulus combination rule in impression formation. *Journal of Experimental Psychology*, *70*(4), 394-400.

Anderson, N. H. (1981). *Foundation of information integration theory*. New York: Academic Press.

Anderson, N. H. (1982). *Methods of information integration theory*. New York: Academic Press.

Anderson, N. H., & Shanteau, J. C. (1970). Information integration in risky decision making. *Journal of Experimental Psychology*, *84*(3), 441-451.

Anderson, N. H., & Weiss, D. J. (1971) Test of a multiplying model for estimated area of rectangles. *American Journal of Psychology, 84*, 543-548.

Bae, G. Y., Choi, J. M., Cho, Y. S., Proctor, R. W. (2009). Transfer of magnitude and spatial mappings to the SNARC effect for parity judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition;*35(6):1506-21. doi: 10.1037/a0017257.

Binet, A. (1969). Children's perceptions. In R. H. Pollack & M. J. Brenner (Trans., Eds.), *The experimental psychology of Alfred Binet* (pp. 93-126). New York: Springer. (Original work published 1890).

Brannon, E.M., & Roitman, J.D. (2003) Nonverbal representations of time and number in animals and human infants. In *Functional and Neural Mechanisms of Interval Timing* (Meck,W.H., ed.), pp. 143–182, CRC Press.

Brannon, E.M., & Terrace, H. S. (1998). Ordering of the numerosities 1 to 9 by monkeys. *Science, New Series*, 282, 746-749.

Brannon, E.M., & Terrace, H.S. (2000) Representation of the numerosities 1-9 by rhesus macaques (Macaca mulatta). *Journal of Experimental Psychology: Animal Behavior Processes*. 26(1), 31-49.

Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society B – Biological Sciences*, 364, 1831–1840.

Busemeyer, J. R. (1991). Intuitive statistical estimation. In N. H. Anderson (Ed.), *Contributions to information integration theory. Volume 1: Cognition*. Hillsdale, New Jersey: Lawrence Erlbaum.

Cantlon, J. F., & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans.*Psychological Science*, *17*(5), 401-406

Church, R.M., & Meck, W.H. (1984) The numerical attribute of stimuli. In *Animal Cognition* (Roitblat, H.L., Beaver, T.G. and Terrace, H.S., eds), pp. 445–464, Erlbaum

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review*, 8(4), 698-707.

Cuneo, D. O. (1982). Children's judgments of numerical quantity: A new view of early quantification. *Cognitive Psychology, 14,* 13-44.

Daar, M., & Pratt, J. (2008). Digit affect actions: The SNARC effect and response selection. *Cortex*, 44, 400 – 405.

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

Dehaene, S. (2009). Origins of Mathematical Intuitions: The case of arithmetic. *Annals of the New York Academy of Sciences*, *1156*, 232–259.

Dehaene, S., Bossini, S., & Giraux, P. (1993). *The mental representation of parity and number magnitude. Journal of Experimental Psychology: General*, 122, 371–396.

Dehaene, S., Dupoux, E., & Melcher, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. Journal of Experimental Psychology: Human Perception and Performance, 16, 626 – 641.

Dehaene, S., Molko, N., Cohen, L., & Wilson, A. J.(2004). *Arithmetic and the brain. Current Opinion in Neurobiology*, 14, 218–224.

De Marchi, S. (1986). Numerical evaluations of collectivities. (S. C. Masin, Trans.). (Original work published 1890). Retrieved from http://www.psy.unipd.it/~masin/DeMarchi1929.pdf

Deschuyteneer, M., De Rammelaere, S., & Fias, W. (2005). The addition of two-digit numbers: exploring carry versus no-carry problems. *Psychological Science*, 47(1), 74-83.

DeStefano, D., & LeFevre, J.-A. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology*, 16, 353-386.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Science, 8*(7), 307–314.

Fias, W. (2001). Two routes for the processing of verbal numbers: Evidence from the SNARC effect. *Psychological Research,* 65, 250 –259.

Freyd, J. J., & Finke, R. A. (1984). Representational momentum. Journal of Experimental Psychology, 10(1), 126-132.

Frith, C.D., & Frith, U. (1972). The Solitaire Illusion: An illusion of numerosity. Perception and Psychophysics, 11, 409-410.

Gallistell, R.C., & Gellman, R. (2000) Non-verbal numerical cognition: from reals to integers. *Trends in Cognitive Sciences*, 4, 59–65

Ginsburg, N. (1976). Effect of item arrangement on perceived numerosity: Randomness vs regularity. *Perceptual & Motor skills*, 43, 663-668.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "Number Sense": The approximate number system in 3-, 4-, 5-, 6-year-olds and adults. *Developmental Psychology, 44(5)*, 1457-1465.

Hubbard, E. M., Piazza, M., Pinel, P, & Dehaene, S. (2005). Interactions Between Number *and Space in Parietal Cortex. Nature Review Neuroscience.* 6(6), 435-48.

Imbo, I., Vandierendonck, A., & De Rammelaere, S. (2007). The role of working memory in the carry operation of mental arithmetic: Number and value of the carry. *Quarterly Journal of Experimental Psychology*, 60(5), 708-731.

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, 106(3), 1221-1247.

Jordan, K. E., & Brannon, E. M. (2006). Weber's Law influences the numerical representations in rhesus macaques (Macaca mulatta). *Animal Cognition*, 9, 159–172.

Kalaman, D., & Lefevre, J. A. (2007). Working memory demands of exact and approximate addition. *European Journal of Cognitive Psychology*, 18, 187-212.

Knops, A., Viarouge A., & Dehaene, S. (2009). Dynamic representations underlying symbolic and non-symbolic calculation: Evidence from the operational momentum effect. *Attention, Perception & Psychophysics*, 71(4), 803-821.

Krause, F., & Lindemann, O. (2013). Expyriment: A Python library for cognitive and neuroscientific experiments. *Behavior Research Methods*, 46 (2), 416-428.

Krueger, L. E. (1972). Perceived numerosity. *Perception & Psychophysics*, 11, 5-9.

Lindemann, O., Abolafia, J. M., Girardi, G., & Bekkering, H. (2007). Getting a grip on numbers: Numerical magnitude priming in object grasping. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 1400 –1409.

Lindemann, O., & Tira M. (2011). Operational momentum in numerosity production judgements of multi-digit number problems. *Journal of Psychology*, 219(1), 50–57.

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16 (1), 149-157. doi: 10.1037/0278-7393.16.1.149

Lourenco, S. F., & Longo, M. R. (2011). Origins and the development of generalized magnitude representation. In S. Dehaene & E. Brannon (Eds.), *Space, time, and number in the Brain: Searching for the foundations of mathematical thought (pp. 225-244).* London, England: Elsevier.

Mandler, G., & Shebo, B. (1982). Subitizing: An analysis of its component process. *Journal of Experimental Psychology: General*, 111, 1-22.

McCrink, K., Dehaene, S., & Dehaene-Lambertz, G. (2007). Moving along the number line: Operational momentum in nonsymbolic arithmetic. Perception & Psychophysics, 69(8), 1324-1333.

McCrink, K., & Wynn, K. (2009) Operational momentum in large-number addition and subtraction by 9-month-old infants. Journal of Experimental Child Psychology, 104, 400-408.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgement of inequality. Nature, 215, 1519–1520.

Myers, L., & Broyles, S. T. (2000). Regression coefficient analysis for correlated binomial outcomes. Journal of

Applied Statistics , 27 (2), 217-234. doi:10.1080/02664760021754

Norman, K. L. (1976). A solution for weights and scale values in functional measurement. Psychological Review, 83(1), 80-84.

Piazza, M., Izard, V., Pinel, P., LeBihan, D., & Dehaene, S. (2004). *Tuning curves for approximate numerosity in the human parietal cortex. Neuron*, 44(3), 547-555.

Pinel, P., Piazza, M., Le Bihan, D., & Dehaene, S. (2004). Distributed and overlapping representations of number, size, and luminance in parietal cortex during comparative judgments. Neuron, 41, 983–993.

Pinhas, M., & Fischer, M. H. (2008). Mental movements without magnitude? A study of spatial biases in symbolic arithmetic. Cognition, 109(3), 408-415.

Pearson, K. (1895) Notes on regression and inheritance in the case of two parents. Proceedings of the Royal Society of London, 58: 240–242.

R Core Team (2013). R: A language and environment for statistical

computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rossetti, Y., & Pisella, L. (2002). Several ‘vision for action’ systems: a geode to dissociating and integrating dorsal and ventral functions. In *Attention & Performance XIX: Common Mechanisms in Perception and Action* (Prinz, W. and Hommel, B., eds), pp. 375–396, Oxford University Press.

Saltzman, I.J., & Garner, W.R. (1948). Reaction time as a measure of span of attention. *The Journal of Psychology*, vol. 25, pp. 227 –241, doi:10.1080/00223980.1948.9917373, PMID 18907281.

Schneider, W., Eschman, A., and Zuccolotto, A. (2012). E-Prime User's Guide. Pittsburgh: Psychology Software Tools, Inc.

Schutz, R. W., & Roy, E. A. (1973). Absolute error: The devil in disguise. Journal of Motor Behavior, 5, 141-153.

Seber, G. A. F., & Wild, C. J. (2003). Nonlinear Regression. New York: Wiley-Interscience.

Shanteau, J., Pringle, L. R., & Andrews, J. A (2007). Why functional measurement is (still) better than conjoint measurement: judgment of numerosity by children and adolescents. *Teorie & Modelli,* 12, 199-210.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: evidence formultiple representations of numerical quantity. Psychological Science, 14(3), 237-243.

Trick, L., & Pylyshyn, Z. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. Journal of Experimental Psychology: Human Perception and Performance, 19, 331-351.

Vidotto G. (2013). Note on differential weight averaging models in functional measurement. *Quality & Quantity*, *47*(2), 811-816.

Vidotto, G., Massidda, D., & Noventa, S. (2010). Averaging models: Parameters estimation with the R-Average procedure. *Psicòlogica*, *31*(3), 461-475.

Vidotto, G., & Vicentini, M. (2007). A general method for parameter estimation of averaging models. *Teorie & Modelli*, *12*(1-2), 211-221.

Vierck, E., & Kiesel, A. (2010). Congruency effects between number mangitude and response force. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36 (1), 204-209.

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, 7, 483– 488.

Wang, M., & Yang, J. (1998). A multi-criterion experimental comparison of three multi-attribute weight measurement methods. *Journal of Multicriteria Decision Analysis*, *7*(6), 340-350.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Non-verbal counting in humans:

The psychophysics of number representation. *Psychological Science*, 10, 130-137.

Wynn, K. (1990). Children's understanding of counting. Cognition, 36, 155-193.

Xuan, B., Zhang, D., He, S., & Chen, X. (2007). Larger stimuli are judged to last longer. *Journal of Vision*, 7, 1–5.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, 74, B1–B11.

Zalinski, J., & Anderson, N. H. (1989). Measurement of importance in multi-attribute models. In J. B. Sidowski (Ed.), *Conditioning, cognition, and methodology. Contemporary issues in experimental psychology* (pp. 177-215). Lanham, Mariland: University press of America.

Zalinski, J., & Anderson, N. H. (1991). Parameter estimation for averaging theory. In N. H. Anderson (Ed.), *Contributions to information integration theory. Volume I: Cognition* (pp. 353-394). Hillsdale, New Jersey: Lawrence Erlbaum Associates.