Universidade do Minho
Escola de Psicologia

Eugénia Alexandra Sabino da Silva Fernandes

Mapping Numbers onto Space in Preschool Children and Adults

outubro de 2014

Eugénia Alexandra Sabino da Silva Fernandes **Mapping Numbers onto Space in Preschool Children and Adults**

UMinho | 2014

Universidade do Minho
Escola de Psicologia

Eugénia Alexandra Sabino da Silva Fernandes

Mapping Numbers onto Space in Preschool Children and Adults

Tese de Doutoramento em Psicologia
Especialidade em Psicologia Experimental e Ciências Cognitivas

Trabalho efetuado sob a orientação do
Professor Doutor Armando Domingos Batista Machado
e do
Professor Doutor François Jacques Tonneau

outubro de 2014

STATEMENT OF INTEGRITY

I hereby declare having conducted my thesis with integrity. I confirm that I have not used plagiarism or any form of falsification of results in the process of the thesis elaboration.
I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

University of Minho, October 27th, 2014

Eugénia Alexandra Sabino da Silva Fernandes

Signature:

Acknowledgments

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

Comecei por conhecer a Psicologia Experimental em 2004, no 1.º semestre do 1.º ano da Licenciatura em Psicologia da Universidade do Minho, quando frequentei a cadeira de Métodos Quantitativos que era leccionada pelo Professor Armando Machado. Como meu orientador no Mestrado e no Doutoramento, o Professor Armando determinou a minha formação enquanto psicóloga e investigadora. De entre as várias qualidades que fazem de si um cientista excepcional, agradeço-lhe especialmente o seu esforço em instigar nos seus alunos um espírito crítico, o incentivar-nos a aprender sempre mais e em tornar-nos investigadores autónomos. Obrigada por esta caminhada.

Agradeço ao Professor François Tonneau pela ajuda no planeamento das experiências e na análise de dados. Mas acima de tudo, agradeço-lhe a sua atitude franca e o seu apoio e conselhos.

Obrigada aos professores e colegas do Laboratório de Aprendizagem e Comportamento Animal pela camaradagem e por me ajudarem sempre que precisei. Obrigada Inês e Carlos pelas muitas vezes em que tal aconteceu. Um especial obrigado à Andréia Kroger-Costa, minha amiga muito querida. Obrigada pelas discussões sobre artigos e planeamento de experiências.

Agradeço à Professora Sylvie Droit-Volet e aos membros do LAPSCO o maravilhoso acolhimento na Universidade Blaise-Pascal. Agradeço também as aprendizagens e discussões académicas e a oportunidade de colectar dados e participar em vários encontros científicos.

A minha mais profunda gratidão à Professora Annie Petrosyan por tão amavelmente ter corrigido o Inglês desta tese, pela amizade e pelos desafios e reflexões resultantes das nossas conversas.

Agradeço à Professora Camila Domeniconi o inestimável encorajamento para iniciarmos o estudo com crianças que veio a tornar-se a pedra angular desta dissertação.

Agradeço à Dr.ª Olívia Mendes por autorizar as colectas de dados no Jardim de Infância de Martim. As colectas no JI foram dos momentos mais cansativos mas também mais reforçadores deste Doutoramento. Agradeço às crianças e aos pais destas, que aceitaram participar nas experiências e trouxeram aos meus dias de trabalho muita doçura e alegria. Quero também agradecer às professoras e técnicas do JI de Martim pela bondade com que me acolheram e incluíram nas suas rotinas. Agradeço em especial à Professora Manuela Marques, à Professora Armanda, à Marta, à D.ª Céu e à D.ª Conceição.

Agradeço aos professores e funcionários da Escola de Psicologia da Universidade do Minho pela minha formação académica e pela sua contínua disponibilidade quando a eles recorri em vários momentos de aflição (dúvidas, requerimentos, ajuda com equipamento, entre outros).

Agradeço à Fundação para a Ciência e a Tecnologia, pela bolsa de doutoramento (SFRH/BD/ 64291/2009) que foi crucial ao desenvolvimento dos estudos nesta Dissertação.

Um grande abraço para a Sara, a Patrícia, o Jorge e os colegas de Doutoramento que comiseraram comigo nas dificuldades e se alegraram comigo nos sucessos. Agradeço especialmente à Rosana, minha irmã de coração.

Obrigada aos amigos de sempre que perdoaram as minhas ausências e sempre me encorajaram.

Agradeço ao meu amigo Sr. Padre Aurélio Araújo Ribeiro, pelo discernimento que me traz ouvir as suas palavras e observar a sua conduta recta, humilde e compassiva.

Em todos os momentos de desânimo ou desespero pude contar com estas pessoas e outras tantas que não tenho espaço para referir. Agradeço-lhes de todo o coração.

Que palavras são suficientes para honrar quem é tudo, aqueles que mais amo nesta vida?

Dedico esta tese de Doutoramento à minha família, os Sabinos. Para mim vocês são a casa e a rua, o partir e o ficar e são o significado de “*só o amor e a morte podem mudar tudo*” [1].

Com muito amor e gratidão, agradeço aos meus pais, Armindo e Lúcia, e à minha querida irmã Ana.

Acredito que tudo devo a Ti, Senhor. Agradeço-Te por ter mantido a esperança.

[1] Khalil Gibran (1926). *Sand and foam*.

Title

Mapping Numbers onto Space in Preschool Children and Adults

Abstract

This dissertation studied how preschool children and adults assign numerosities onto a spatial continuum. In Number-to-Position (NTP) tasks, participants are presented a horizontal rectangular response bar. They are trained to select the bar's left endpoint when they see a minimum numerosity (e.g., '1') and to select the bar's right endpoint when they see a maximum numerosity (e.g., '10'). Next, they are tested also with intermediate numerosities and have to estimate their location along the 1-to-10 bar. Previous symbolic NTP studies (Arabic digits or spoken words) have shown that preschoolers produce a linear-like response pattern in small numerical ranges (1-to-10), with constant spacing along a line for consecutive numbers. When the range is 1-to-100 or 1-to-1000, responses for the larger numbers are compressed at the right portion of the line, in a logarithmic-like pattern. Also, a developmental log-to-linear shift occurs across school age groups. Some authors have proposed that number and space are inherently associated, so that numbers are spontaneously mapped onto space in a logarithmic scale, which may be linearized with schooling (*mental number line* hypothesis; Dehaene et al., 2008).

In Study 1 we tested preschoolers and adults in NTP tasks with symbolic (spoken words) and nonsymbolic (arrays of dots, sequences of tones) numerical conditions. Although both groups' average location curves increased with numerosity, inspection of individual single-trial scatterplots revealed that, contrary to adults, preschoolers' smooth and increasing curves were an averaging artifact. Instead of responding along the line's extent (continuous pattern), many preschoolers restricted their responses to the endpoints (bi-categorical pattern) or to the middle and the endpoints (tri-categorical pattern). Subsequent studies investigated the effects of three pre-training histories on NTP performance with arrays of 1-to-9 and 10-to-90 dots. In Study 2, prior to NTP testing, participants learned to respond along the bar as a function of increasingly darker stimuli (Brightness-to-Position). Study 3 isolated the "mechanical" component of spatial responding, with participants having to select different bar locations as a function of cartoon

images (Figures-to-Position). Finally, in Study 4 participants received a perceptual training on discrimination within the numerical range tested in the NTP task.

The majority of preschoolers tested solely in NTP tasks responded categorically. Most important, group average curves were misleading and did not represent individual performance. Moreover, only the Brightness-to-position pre-training (Study 2) significantly improved preschoolers' use of the response bar in the NTP task. For these reasons, our results challenge both the assumptions that (i) mapping numbers onto space is an innate intuition; and that (ii) NTP responses directly mirror a *mental number line*. These findings should urge researchers to focus on individual performance, even in previous symbolic NTP studies because their main or sole focus of analysis has been the average group curve.

Título

Mapeamento de Números no Espaço em Crianças Pré-escolares e Adultos

Resumo

Esta dissertação estudou como crianças pré-escolares e adultos distribuem numerosidades num continuum espacial. Em tarefas Número-para-Posição (NPP), aos participantes é apresentada uma barra de respostas rectangular na horizontal. Os participantes são treinados a seleccionar o extremo esquerdo da barra quando vêem uma numerosidade mínima (e.g., '1') e a seleccionar o extremo direito da barra quando vêem uma numerosidade máxima (e.g., '10'). Em seguida, são testados com numerosidades intermédias e têm de estimar a posição destas ao longo da barra que vai de 1 até 10. Estudos anteriores de NPP simbólicos (dígitos Árabes ou palavras faladas) têm mostrado que crianças pré-escolares produzem um padrão aproximadamente linear em intervalos numéricos pequenos (1-até-10), com um espaçamento constante ao longo da linha para números consecutivos. Quando o intervalo é 1-até-100 ou 1-até-1000, as respostas para os números maiores estão comprimidas na porção direita da linha, num padrão aproximadamente logarítmico. Para além disso, existe uma transição desenvolvimental de log-para-linear em função da idade de escolaridade. Alguns autores têm proposto que número e espaço estão inerentemente associados, a ponto de os números serem mapeados espontaneamente no espaço numa escala logarítmica, que pode ser linearizada com a escolaridade (hipótese linha numérica mental, Dehaene et al., 2008).

No Estudo 1 testámos crianças pré-escolares e adultos em tarefas NPP com condições numéricas simbólicas (palavras faladas) e não-simbólicas (conjuntos de pontos, sequências de sons). Embora as curvas médias de localização de resposta de ambos os grupos aumentassem com a numerosidade, a verificação dos gráficos individuais com respostas ensaio-a-ensaio revelou que, ao contrário dos adultos, as curvas contínuas e crescentes das crianças eram um artefacto resultante da média. Ao invés de responderem na extensão da barra (padrão contínuo), muitas crianças restringiram as suas respostas às duas extremidades (padrão bi-categorial) ou às duas extremidades mais a posição média (padrão tri-categorial). Os estudos que se seguiram investigaram os efeitos de três histórias de pré-treino no desempenho em NPP com conjuntos de 1-até-9 e 10-até-90 pontos.

No Estudo 2, antes do teste em NPP os participantes aprenderam a responder ao longo da barra em função de estímulos progressivamente mais escuros (Luminosidade-para-Posição). O Estudo 3 isolou a componente “mecânica” do responder espacial, com os participantes a terem de seleccionar diferentes localizações na barra em função de imagens de desenhos animados (Figuras-para-Posição). Por último, no Estudo 4 os participantes receberam um treino perceptual em discriminação no intervalo numérico testado na tarefa de NPP.

A maioria das crianças testadas apenas em tarefas NPP respondeu categorialmente. Mais importante, as curvas médias de grupo eram enganosas e não representavam o desempenho individual. Para além disso, apenas o pré-treino de Luminosidade-para-Posição (Estudo 2) melhorou significativamente o responder na barra das crianças durante a tarefa NPP. Por estas razões, os nossos resultados desafiam os pressupostos de que (i) o mapeamento de números no espaço é uma intuição inata; e de que (ii) as respostas de NPP espelham directamente a *linha numérica mental*. Estes resultados devem encorajar os investigadores a focarem-se no desempenho individual, mesmo em anteriores estudos de NPP simbólicos, uma vez que o seu maior ou único foco de análise tem sido a curva média do grupo.

Table of Contents

Abbreviations, acronyms and symbols

ANOVA	Analysis of variance
ITI	Intertrial Interval
M	Mean
Med	Median
SD	Standard Deviation
SEM	Standard Error of the Mean
Lin	Linear
Log	Logarithmic
n.s.	Not significant
RGB	Red, Green, Blue color system
OLS	Ordinary Least Squares
2AFC task	Two-alternative forced-choice task
Cm	centimeter
In	inches
Ms	millisecond
S	second
MTS	Matching To Sample
RTs	Reaction Times
ANS	Approximate number system
MNL	Mental Number Line
BP	Bisection Point
WR	Weber Ratio

Figures

Chapter I: General Introduction

Chapter II: STUDY 1: Preschoolers' categorical vs. adults' continuous mapping patterns in symbolic and nonsymbolic number-to-position tasks

Chapter III: STUDY 2: Control of a response continuum by the numerical stimulus continuum: the effects of pre-training on a non-numerical continuum

Chapter IV: STUDY 3: Control of a response continuum by the numerical stimulus continuum: isolating the effect of training responses in different locations of the bar

Chapter V: STUDY 4: Control of a response continuum by the numerical stimulus continuum: isolating the effects of a perceptual training on Number-to-Position performance

Chapter VI: GENERAL DISCUSSION

Tables

Chapter III: STUDY 2: Control of a response continuum by the numerical stimulus continuum: the effects of pre-training on a non-numerical continuum

Chapter IV: STUDY 3: Control of a response continuum by the numerical stimulus continuum: isolating the effect of training responses in different locations of the bar

Chapter V: STUDY 4: Control of a response continuum by the numerical stimulus continuum: isolating the effects of a perceptual training on Number-to-Position performance

CHAPTER I : GENERAL INTRODUCTION

1. Number: the "poor cousin" in the Space, Time and Number triad?

In our everyday life, it is usual for us to experience a strong positive correlation between the three stimulus properties: spatial extension, duration and numerosity. For instance, the larger the number of elements that form a set (e.g., the number of students sitting in class) the larger the space it occupies (e.g., larger length, area, volume). Similarly, in situations where a numerical sample is presented sequentially (e.g., number of consecutive drumbeats), it is usual to experience that the more elements, the larger the sample's total duration.

Similar correlations are also experienced by nonhuman animals' in everyday situations. Consider, for example, the case of a bird in a cage with two feeders. Assuming that the seeds are approximately of the same size, a hungry bird will probably select the feeder which has the larger number of seeds, which is also the feeder with larger volume of food. In another example, a common feeding situation with captive elephants, their trainer cuts apple pieces into a basket. Even if the apple pieces are hidden once inside the basket, the more that are placed, the longer this action will last. Given the opportunity to select between two different quantities of food, elephants will select the basket associated with larger numerosity, total amount and longer duration (Irie-Sugimoto, Kobayashi, Sato, & Hasegawa, 2009). The examples are not restricted to animals in captivity, nor to animals that often interact with humans. In their natural habitat, during many situations of inter-group collaboration, conflict, parental investment and predator avoidance, judgments of relative quantity seem to predict the animals' behavior (Honig & Stewart, 1989; Gómez-Laplaza, & Gerlai, 2010; Krebs, 1978; Stevens, Wood, & Hauser, 2007).

In the preceding examples, one is not able to determine which particular stimulus property controls behavior. But within the number, space, and time triad, number has always been regarded as the "poor cousin" (Emmerton & Renner, 2009; Davis & Memmott, 1982; Davis & Pérusse, 1988). One question that for a long time divided researchers was whether preverbal children and animals could discriminate the purely numerical features in the environment, or if overall amount (spatial extent) or duration must co-vary (Shettleworth, 2010, p. 341; Piaget, 1952; Gelman & Gallistel, 1978). Until recently, in the animal and nonverbal human cognition literature, even researchers studying numerical abilities viewed number as a "last resort" cue (Davis & Memmott, 1982; Davis & Pérusse, 1988; Davis, 1993; Seron &

Pesenti, 2001; Clearfiled & Mix, 1999, 2001). A long lasting notion was that, if available, other continuous variables such as area, length, or duration, would be more salient to animals and preverbal children than numerical information.

Such reluctance in considering number as elementary a stimulus' property as, say, duration or length might have occurred because of the reduced scientific interest and belief in animal numerical abilities in the aftermath of the famous Clever Hans incident. At the beginning of the 20th century, Wilhelm von Hosten, a retired German professor and animal trainer, claimed to have taught his horse Hans to, among several other impressive feats, solve arithmetic summation, multiplication and division problems. In public displays of Hans's prowess, von Hosten would invite audience members to verbally present or draw on a board simple arithmetic problems, to which Hans responded by tapping its hoof a number of times indicating his answer. The accurate performance of the horse gained such attention that a scientific committee was appointed to investigate the horse's ability to think, count or solve arithmetic problems. The verdict, however, revealed inconclusive and no statements were made regarding the horse's cognitive abilities. It was simply acknowledged that von Hosten was not intentionally deceiving everyone (Stumpf, in Pfungst, 1911). A second investigation of Hans' abilities was later carried out by psychologist Oskar Pfungst, the research assistant of psychologist Carl Stumpf, who had led the first committee enquiry. Pfungst operated under the hypothesis that the horse had no mathematical ability and that its behavior was controlled by cues given by the trainer or the audience. He implemented tests in which the arithmetic problems presented to the horse would occasionally differ from the ones presented to the trainer and found that the horse would invariably tap its hoof according to the trainer's problem. After systematic testing, it was concluded that the horse could detect when to stop tapping its hoof on the basis of unintentional body cues, such as slight movements of the head or eyebrow and changes in posture of the trainer or the audience members.

In the long-term, the clever Hans incident proved to be an influential event for Animal Cognition research. Nowadays, it is agreed upon that human and animal psychological studies demand precise, rigorous experimental designs and a setting of controls against experimenter bias (Roberts, 1998, p.9). But despite this beneficial long-term influence, the immediate effect of the Clever Hans incident was a general skepticism towards other "clever" animal reports and the scientific field of Animal Cognition, animal numerical abilities in particular (Davis & Memmot, 1982). By the early 20th century, with the rising influence of John Watson's behaviorist view of the philosophy of psychological science, American psychologists were disinclined to study the so-called higher mental activities in animals (Watson, 1914, cit. by Davis & Memmot, 1982).

However, during the second half of the 20th century, an increasing number of laboratory studies led by psychologists, demonstrated that animal's behavior could be controlled by the number of their own emitted responses (Mechner, 1958; Rilling & McDiarmid, 1965; Fetterman, 1993; Machado & Rodrigues, 2007) or by the number of exteroceptive events, such as light flashes, tones, shocks or food reinforcers (Seligman & Meyer, 1970; Thomas, Fowlkes, & Vickery, 1980; Fernandes & Church, 1982; Davis, 1984; Alsop & Honig, 1991; Pepperberg, 1994; Roberts, Macuda, & Brodbeck, 1995).

An important endeavor in the experimental study of numerical discrimination has been the disentanglement of the "number" property from concomitant temporal and spatial cues. Not surprisingly, the controls implemented to study purely numerical stimuli depend on the sensory modality. For example, when presenting subjects exteroceptive visual stimuli, such as arrays of dots, the numerical samples may be matched on overall brightness, density, line length, contour, area, and homogeneity of the items that constitute the set (Jordan & Brannon, 2006b, 2006c, Jordan, Brannon, Logothetis, & Ghazanfar, 2005; Emmerton & Renner, 2006). When numerical samples are presented in the auditory modality, such as sequences of tones, samples may be equalized in terms of total duration, delay between the onsets of two consecutive events (tempo), average root mean square (rms) power, and amplitude (Jordan & Brannon, 2006a; Jordan Maclean, & Brannon, 2008; Hauser, Dehaene, Dehaene-Lambertz & Patalano, 2002). Additionally, experimenters may opt to systematically vary fundamental frequency, intensity, pitch, and timbre (Hauser, Dehaene, Dehaene-Lambertz & Patalano, 2002)

Putting aside the debate on the relative salience of each stimulus dimension (e.g., Mix, Levine, & Huttenlocher, 1997), what is certain is that when these controls are implemented in nonverbal tests of sensitivity to number (Shettleworth, 2010, p. 430) and applied to a wide range of nonhuman and human organisms, they can respond to the purely numerical features in the environment (Emmerton & Renner, 2009; Capaldi & Miller, 1988; Cantlon & Brannon, 2007). In addition, number discrimination findings converge with the studies that have isolated "pure" features of both time and space. For example, discrimination of numerosities, durations, and size are both (i) ratio dependent and (ii) obey Weber-Fechner's law or a generalized version of it (Brannon, 2006; Whalen, Gallistel, & Gelman, 1999; Roberts, 1995, 2005, 2006; Jordan & Brannon, 2006c; Droit-Volet, Clément, & Fayol, 2003; Droit-Volet, 2010; Dehaene, Dehaene-Lambertz, & Cohen, 1998; Machado & Rodrigues, 2007). The former property signifies that (i) it is not the absolute difference which determines the degree of discrimination between two samples, but their ratio. For example, consider the situation when an organism must discriminate between 4 and 8 dots or between 8 and 12 dots. In both discriminations, the difference between the numerosities is 4, but because the ratio is larger in the first than in the second discrimination (1:2 against 2:3, respectively), the first discrimination is easier. Conversely, when a ratio is held constant across several pairs of values, discriminability is approximately equal (e.g., 4 vs. 8, 8 vs. 16, 2 vs. 4). The second property of number, space and time discrimination is observed because, (ii) as stimulus' magnitude increases, so does the variability in responding. More precisely, the standard deviation of numerical "estimates" is proportional to the mean of these "estimates".

In a complementary approach, other studies have demonstrated that the systematic manipulation of the magnitude of one of the stimulus dimensions may affect the perception of the other dimensions (Ginsburg, 1976, 1978, 1980; Ginsburg & Goldstein, 1987; Foster, 1978; Frith & Frith, 1972; Howe & Jung, 1987; Jenkins & Cole, 1982; Fischer, 2001; Emmerton & Renner, 2006; Cohen-Kadosh, Lammertyn, & Izard, 2008; Haun, Jordan, Vallortigara, & Clayton, 2010; Beran et al., 2008, 2011; Lourenco & Longo, 2010; Roberts & Mitchell, 1994). As a result, researchers in the fields of both Cognitive Neurosciences (e.g., Dehaene, 2003; Walsh, 2003; Dehaene & Brannon, 2010a, 2010b, 2011), and behavioral Psychology and Animal Cognition (e.g., Meck & Church, 1983; Meck, Church, & Gibbon, 1985; Whalen, Gallistel, & Gelman, 1999; Roberts, Coughlin, & Roberts, 2000), have proposed that a generalized sense of magnitude underlies the notions of size, duration and numerosity.

The current dissertation will discuss the data obtained with a procedure which, according to previous authors who have implemented it, demonstrates how "the mapping of numbers onto space is a universal intuition" (Dehaene et al., 2008). But before we present the procedure, we will first describe an interesting phenomenon of number-space association in human adults.

2. The association between number and space: Galton's Number Forms

At the end of the nineteenth century, Francis Galton (1822-1911) had dedicated himself to the study of mental imagery of humans or, in his own words, of people who are "apt to think in visual images; not in fancied words, nor in an abstract manner" (1881; Galton, 1880b). These people who possessed vivid mental presentations were able to imagine a series of pictures, vivid in colour and well defined in form, so that in many cases they appeared external to them.

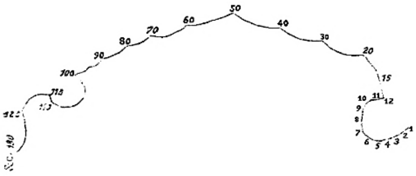
Both in a paper published in Nature (1880a) and a memoir read before the Anthropological Institute (later published in the Institute's Transactions, in 1881), Galton first described the ability of some people to think of numerals in visual imagery. Galton estimated that about 1 in 30 adult men and 1 in 15 women invariably associated numbers to a definite "pattern or form" (1880a). Although Galton himself saw no "Number Form", he collected reports of 80 people who, whenever they thought about numbers, "visualised numerals in diagrammatic and colored shapes" (1880a).

To illustrate, in Figure 1 we present six exemplars of Number Forms[2]. As displayed in Figure 1, Number Forms varied markedly across subjects. They could "consist of a mere line of any shape, of a peculiarly arranged row of figures, or of a shaded space" (Galton, 1881). Even though the reported Number Forms varied from a simple line to more elaborate figures in terms of colour and depth, they were all characterized by their vividness and by being automatically activated. As the subjects put it, "I cannot think of any number I at once result in its peculiar place" (Bidder, Fig. 1) and "in thinking of a number, it always takes its place in the figure" (D.A., Fig. 1) (1880a). This figure was invariable, for each subject' s numbers "show themselves in a definite pattern that always occupies an identical position in their [the subjects'] field of view, with respect to the direction in which they were looking" (Galton, 1883).

[2] We direct the interested reader to Galton's 1881 paper and the "Number-Forms" chapter in his 1883 book, for several of these diagrams. They are accompanied by descriptions, as well as excerpts from the correspondence between Galton and the subjects who reported the Number Forms.

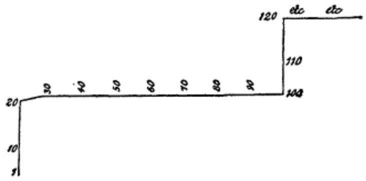
Galton (1880a)

Bidder



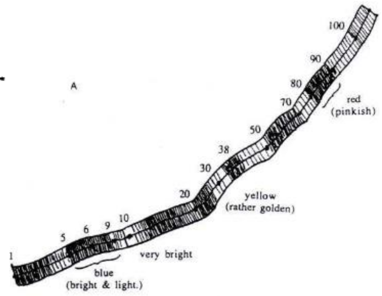
"*Every number (...) is always thought of by me in its own definite place in the series, where it has (...) a home and an individuality*."

J.S.



"Figures present themselves to me in lines. They are black on a white ground. There is no light or shade, and the picture is invariable."

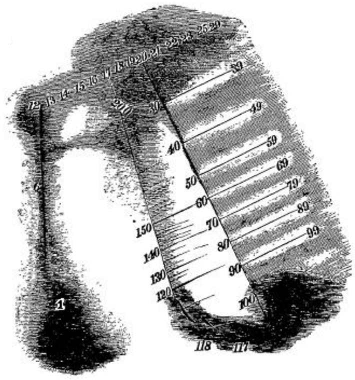
Fig.4



[A colored ribbon, undulating rightward. Small numbers are in the bluish segment; as numbers become larger, the ribbon is yellow, then red/pinkish]

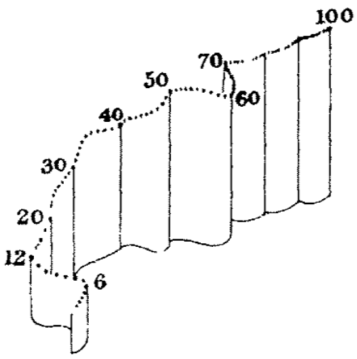
Galton (1881)

D.A.



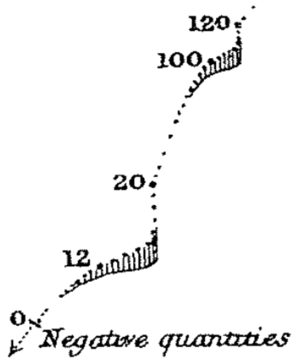
"From the very first I have seen numerals up to nearly 200 range themselves always in this particular manner, and in thinking of a number it always takes its place in the figure."

R.N.



[The base and the verticals are merely to explain the perspective]

C.H.



[The figures are on a path that lies over undulations, and is seen obliquely]

Figure 1: Diagrams illustrating six subjects' "Number Forms", as reported in Galton's 1880a and 1881 papers on "Visualised Numerals".

Another common property was that the spatial representation always preserved the ordinal arrangement between consecutive integers. Subjects would report that "numbers appear beyond one another, stretching away into space" or that "the numbers (…) are associated with points on an ascending and descending scale"

(1880a). Moreover, small integers tended to present themselves continuously and there would be occasional discontinuities or abrupt changes at the decade transitions (from 19 to 20, 29 to 30 and so on). Lastly, the run of the lines for smaller numbers was distinct and rarely analogous to that of larger numbers. As numbers increased, the Number Form became fuzzier..

What could be the basis for these Number Forms? When questioned, subjects answered that they had always thought about numbers in that way and were unable to determine the origin of their Number Form. Thus, they were not explicitly taught and were present before the subject learned to read. Galton hypothesized that Number Forms were "survivals of the earliest of (…) mental processes, and a clue to much that is individual in the constitution of the mind (1880a)". He proposed that Number Forms resulted from an innate ability to represent numbers in an orderly fashion. Galton suggested that the adults' Number Forms probably started as "mnemonic diagrams, invented by children when they were learning to count *verbally*", so that "the *sounds* of the successive numerals [started] being associated with the successive points of the form. Also, when the children began to read, the visual symbols of the numerals quickly supplanted the verbal ones, and established themselves permanently in their place" (Galton, 1880a). In other words, the visual symbols (e.g., Arabic digits) overrode the early verbal ones (words). That some Number Forms' shapes are uneven would be explained by "personal fancies each person had for certain shapes", for similar reasons as to why handwriting styles are personalized.

About half a century later, Spalding and Zangwill (1950) reported the clinical case study of patient A.L.' s marked calculation impairment following a gunshot wound. A metal object had entered A.L.'s "left parieto-occipital region and crossing the mid-line to come to rest ½ in. [1.27 cm] above the right petrous temporal bone". As a result, A.L. lost his ability to name objects (nominal dysphasia / anomic aphasia) but was still able to understand their purpose. He presented complete dyslexia but would still comprehend speech. In addition, he did not perceive the right half of each visual field (right homonymous hemianopia). Some months after the injury, A.L. experienced some improvement in his speech capacity but whenever he was out of his home, he would frequently lose his way. His everyday functioning was altered, and because he had a clear insight into his disabilities he would mostly keep to his home and avoid people who were not family members.

Five years after the lesion, in addition to personality changes, A.L.'s major disability was a defect in visual memory. Ever since his injury, he could not copy and draw from memory and his topographical sense, of where things and houses stood in relation to each other, had become "twisted". Even when given verbal directions, A.L. could only remember up to two changes in direction. Yet another major impairment was A.L.'s marked difficulties in performing simple arithmetic problems (e.g., 6 + 5). On one occasion when he was completing an arithmetic screening test, he spontaneously remarked that he "used to have a plan of numbers, but had lost it", or, more accurately, ever since the injury it was no longer distinct. This plan had previously been useful when performing calculation, but because it was now difficult to hold it in mind, it had become "more hindrance than help" (Spalding & Zangwill, 1950).

When asked to draw his plan, A.L. made a diagram up to number 12, and had to instruct others to continue the number line. Later, he himself drew a complete diagram, which corresponded accurately to the first one. A.L. also reported that before the injury he had used "forms" for other measures and he was still able to draw his forms of months, days of the week and the alphabet. Spalding and Zangwill recognized that A.L.'s previous stereotyped mode of visualizing numerals matched the Number Forms reported earlier by Galton (1883). Namely, as in Galton's observations on Number Forms, A.L.'s "plan of numbers" "ran upward rather than downward" and they "extended into the third dimension" (Spalding & Zangwill, 1950).

Number Forms similar to the ones reported in Galton's studies were also found in the 1992's study by Seron and colleagues (Seron, Pesenti, Nöel, Deloche, & Cornet, 1992). The authors passed a short questionnaire to French psychology students (total sample = 194, females = 153) to screen for possible Number Forms. Among the 49 students who reported a Number Form, 26 completed a second, extended questionnaire. This questionnaire comprised Yes/No/Don't' know responses, open responses and frequency-of-use judgments. It covered many aspects of Number Forms; their shape, color, projection in 3-D space, how automatically they were activated, if they were affected by the environment and the subject's states, if they were used when performing calculations, etc. In Seron and colleagues's study, most number representations were continuous lines, scales or grids (52%). A smaller number were cases of coloured codes (26%) and, finally, there were three subjects who reported analogical representations (11%). What is more, the diagrams and verbal descriptions of the continuous lines and scales presented by the university students shared many common features with Galton's Number Forms, such as: (i) appearing automatically and being invariable in time, (ii) having been present early in childhood, (iii) the more the subject focuses on a number, the more vivid its region becomes in detriment of the rest of the line, (iv) numbers were spatially organized according to their ordinal value, and (v) their spatial organization seemed to be influenced by the base-ten organization of the Arabic number system.

The origin of the Number Forms remains unknown. In his book "The Number Sense" (1997), neuroscientist Stanislas Dehane proposed that the occurrence of Number Forms could be a result of "how cortical maps of space and number are formed during development" (p. 85). According to Dehaene's theory of number representation, both humans and non-human animals share a primitive sense of number (1997, p. 40). In this representation system, which is commonly designated as the Approximate Number System (ANS) or the Mental Number Line (MNL), "numerical quantities are represented as inherently variable distributions of activation over an oriented analogical number line obeying Weber-Fechner's law" (Dehaene, 1992; Dehaene et al., 1990; Halberda, Mazzocco, & Feigenson, 2008). Children at about 3-years of age start experiencing formal schooling and because of that the initial number line (i.e., psychological representation) is altered by the new numerical knowledge. Children are trained in a base-ten counting system and learn arithmetic, among other cultural tools of precise measurement. As a result, at this age the amount of cortex dedicated to the number maps, which seems to be situated in the intraparietal sulcus (Simon et al., 2002; Nieder & Miller, 2003, 2004; Piazza et al,, 2004), progressively expands. This expansion of the numeral network may cause it to overlap with surrounding cortical maps, including those coding for visual and spatial stimulus properties such as color, form, and location. Such overlap between cortical areas could explain occurrences of "seeing the color and location of numbers" (Dehaene, 1997, pp. 85-86; Hubbard, Pinel, Piazza, & Dehaene, 2005).

Because of its reliance on introspection, researchers have been reluctant to study Number Forms. Nevertheless, the similarities between the Number Forms observed across different studies (Galton, 1880b, 1881, 1883; Bertillon, 1880, 1881, 1882; Spalding & Zangwill, 1950; Seron et al., 1992), as well as the reliability of the intra-subject reports, seem to discourage the notion that they are inventions made up to meet the experimenters' expectations.

Next, we will address more recent lines of research that have also been pointed out as proof for a special association between number and space perception in humans.

3. Recent findings on the Number-Space Association

The experimental studies relating numerical and spatial discrimination in humans may be divided into three main research lines: the SNARC effect, Line Bisections, and Number-to-Position studies. The acronym SNARC stands for "*Spatial-Numerical Association of Response Codes*", and was coined by Dehaene to describe the finding that, in "more vs. less" judgment tasks, humans' reaction times (RTs) are affected by the distance between the numerical stimuli (Moyer & Landauer, 1967; Buckley & Gilman, 1974) and by the spatial location of the response keys (Dehaene, Dupoux, & Mehler, 1990; Dehaene, Bossini, & Giraux, 1993). The effect of the second variable was discovered serendipitously by Dehaene, Dupoux and Mehler (1990, Experiment 2), when they replicated Hinrichs et al. (1981)'s number comparison study with French students. The task proceeded as follows. On each trial, a target Arabic digit appeared on screen for 2-s. This target digit could be any number between 31 and 99. The participants had to decide whether the trial's target number was smaller or larger than the standard number "65". To that end, participants responded on two response keys. When the target was larger than the standard, half of the participants had to respond with their right hand (larger-right) and the other half with their left hand (larger-left). This spatial organization of the responses influenced participants' RTs. Specifically, when the numerical sample was smaller, they responded faster and more accurately if the correct response key was situated at their left, rather than their right side. Complementarily, the presentation of larger numerical samples expedited responses at the participants' right side (i.e., larger-right). In a later experiment, in which subjects responded with their hands crossed throughout the session, "Larger" responses were faster when participants answered with their left hand at the right side, than with their right hand responding at the left side (Dehaene, Bossini, & Giraux, 1993).

The interest in the SNARC effect was such that, out of the three number-space lines of research, it became the most prolific in terms of number of publications. Nonetheless, that does not prevent it from seeming incongruences across studies and the explanation for the SNARC effect is still a matter of discussion. Particularly, researchers differ as to whether the left-small and large-right associations are related to cultural factors such as direction of reading or writing. The "culture" argument states that the SNARC effect occurs because of adults' extensive exposure to the relation that "when a series of numbers is written, larger numbers appear to the right of smaller numbers" (Dehaene, Dupoux & Mehler, 1990; Van Galen & Reistma, 2008). On one hand, Dehaene and colleagues (1993, Exp. 7)' experiment with Iranian students who had learned to read from right to left, showed that the direction of the SNARC effect depended on the amount of exposure to Western (French) culture (see also Zebian, 2005). On the other hand, however, reading and writing direction do not explain the vertical SNARC effect, i.e., that "Larger" responses are facilitated by a vertical bottom-to-top direction (Ito & Hatta, 2004; Schwarz & Keus, 2004). Additionally, Wood, Nuerk and Willmes (2006a) were not able to replicate the "crossing hands" results obtained by Dehaene and colleagues (1993)' study, which pointed to a possible effect of the experimental manipulations. Indeed, direction of reading cannot explain why even for the same Western participant, the SNARC effect depends on the type of instructions and, more important, why it can be reversed (Bächtold, Baumüller, & Brugger, 1998; Gevers & Lammertyn, 2005; Gevers, Lammertyn, et al., 2006; Fischer, 2006; Notebaert et al., 2006; Wood, Nuerk, & Willmes, 2006b; Müller & Schwarz, 2007; Santens & Gevers, 2008; Fischer, Shaki, & Criuse, 2009).

A second line of evidence for the effect of numerical information on spatial responding comes from Line Bisection studies. In these experiments, the participant is presented with a line segment flanked on each extremity by numerals (Arabic digits) or sets of dots (Figure 2). However, these numerical "anchors" are irrelevant to the participants' task, which consists solely of indicating the line's midpoint (i.e., participants have to bisect the line). Evidence for a number-space association occurs because, when the line is flanked by two digits or two sets of dots, the location that participants select to bisect the line is biased towards the side where the largest numerical anchor is located at (de Hevia & Spelke, 2009, 2010; Longo & Lourenco, 2007).

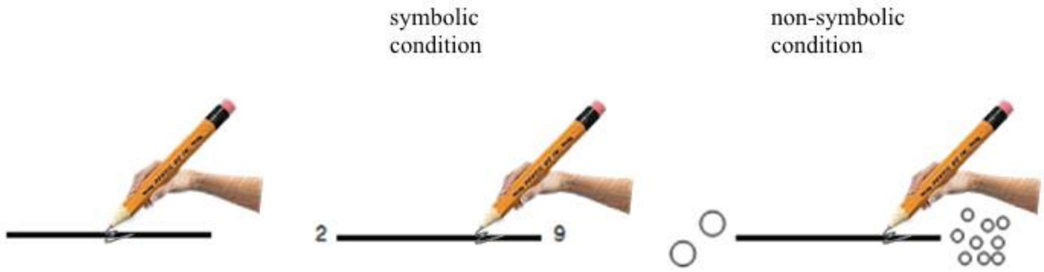


Figure 2. Illustrations of three Line Bisection tasks. The first, leftmost image depicts a standard Line Bisection task. The other two images illustrate the manipulation of the line being flanked by two numerosities (anchors), presented symbolically (Arabic digits) or non-symbolically (sets of dots). They also illustrate the spatial bias, towards the side with the larger numerical anchor.

The third source of evidence for an association between space and number comes from tasks of mapping numbers onto space, also named Number-to-Line or Number-to-Position tasks. Whereas the two previous procedures required a single spatial response, the Number-to-Position tasks require the differentiation of several spatial positions as a function of different sample numerosities. Because the studies we will present on this dissertation were conducted with the Number-to-Position procedure, we will now describe it in greater detail.

4. Number-to-Position findings with Arabic digits

Number-to-position tasks may be considered a particular type of cross-modality matching tasks (Stevens, 1960, 1966), where participants are presented with a numerical sample and, rather than verbally estimating its magnitude, they must translate it into a location along a spatial *continuum* (Siegler & Booth, 2005). Also known as number-to-line tasks, they usually entail learning the correspondence between two numerosities and the two endpoints of a horizontal line. Afterwards, participants are presented with intermediate numerosities and instructed to select their corresponding positions along the line, taking into account the distance between the novel value and the two anchor numerosities.

In Siegler and Opfer (2003)'s landmark study, participants from four schooling ages (second, fourth, sixth graders and undergraduates) were shown a paper sheet with a printed line segment flanked by two Arabic numerals, "0" at the left endpoint and "100" or "1000" at the right endpoint. A third numeral was printed centrally above the line, and represented the numerical sample that should be positioned along the line. Participants were initially taught to mark the leftmost part of the line when the sample was "0" and the rightmost position when it was "100" or "1000", depending on the condition. Additionally, they were instructed that the sample "50" (or "500") corresponded to the midpoint position. Afterwards, they were presented with numerals ranging from the smallest to the largest anchor and were asked to estimate their position in the 0-100 or 0-1000 line segments. Figure 3 shows an example of one trial from the 0-100 condition.

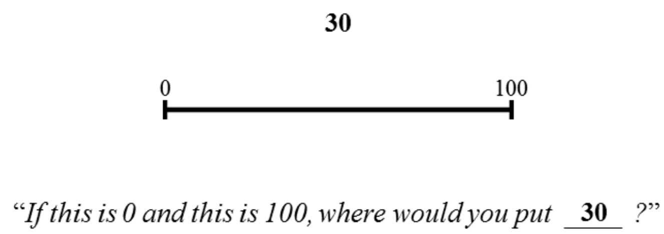


Figure 3. Example of one trial in a 0-100 Number-to-position experiment, with Arabic digits (from Siegler & Opfer, 2003). The participant is presented a line segment, flanked at left by the smallest anchor (0), and at right by the largest anchor (100). The sample is a number, in this case "30", printed centrally above the line. The verbal instruction is as shown in italics.

The authors' main objective was to see how response locations along the spatial *continuum* varied with the numerical sample and age groups. In the 0-1000 number line task, it was found that second and fourth graders presented a logarithmic-like pattern of response locations, with responses for larger numbers compressed at the right portion of the line (Siegler & Opfer, 2003). This result is illustrated at the right half of Figure 4. In the figure, the y-axis (response location) is the median of the relative distance from the line's left endpoint. As such, it can take any value from 0, when the participant selects the leftmost anchor, to 1, when response occurs at the rightmost anchor. Thus, a response location of 0.5 corresponded to the line's midpoint. The data show that median response location increases with numerical sample (Arabic digits).

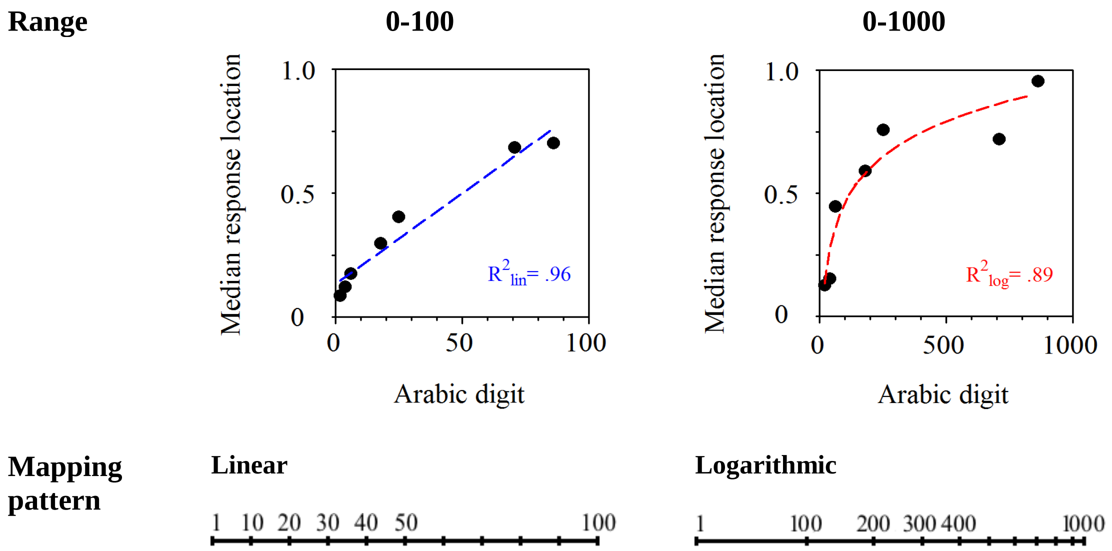


Figure 4. Data from 2nd Graders tested in Number-to-Position tasks, in Siegler & Opfer (2003)'s study. In the 0-100 task, the group's median response locations were better fit by a simple linear function, which is depicted in blue (R^2_{lin} = .96). In the larger range, 0-1000, the curve was better described by the logarithmic function (R^2_{log} = .89), depicted in red. The Figure's lower row represents how "ideal" linear (left) and logarithmic (right) mappings would translate into selected positions along the number-line.

Sixth graders and undergraduates, however, showed a linear pattern in the 0-1000 task, keeping an approximately constant spatial distance between constant numerical intervals. Interestingly, though second graders exhibited a logarithmic pattern in the 0-1000 task, in the 0-100 task they exhibited a linear pattern (see left half of Figure 4).

In a subsequent study with 0-100 number line tasks, which included even younger children, a log-to-linear shift was found between kindergarten and second grade groups (Siegler & Booth, 2004). This linearization of an earlier logarithmic-like pattern of responses across schooling age groups has since been replicated in Arabic digits number-to-position tasks and appears to be quite robust (Booth & Siegler, 2006; Opfer & Siegler, 2007; Thompson & Opfer, 2008; Berteletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010).

5. Are number and space inherently related? The *mental number line* hypothesis

Results from Number-to-position experiments have been compared to the psychophysical numerosity properties found in other number discrimination procedures. As we have previously mentioned, an ubiquitous empirical finding from other procedures is that humans and nonhumans' number discrimination follows Weber's Law. There are two main contending models of psychological scaling that instance this law. The first, consistent with Fechnerian psychophysics, is that the subjective value of a numerosity is a logarithmic function of its physical value (Dehaene, 2003). Hence, when stimuli bear the same physical ratio, the psychological differences are equal. Yet another account that mathematically predicts the same result is that of a linear psychological scale with proportional, scalar variability (Whalen, Gallistel, & Gelman, 1999). Which account holds, and under which experimental condition remains a topic of debate in animal and human numerical cognition (Brannon, 2006; Roberts, 2006; Beran, Johnson-Pynn, & Ready, 2008; Merten & Nieder, 2009).

Such inquiry on the format of the subjective scale also extends to the interpretation of number-to-line findings. A matter of debate is how directly do responses reveal the form of the transformation of the physical intensity into a psychological magnitude (Shepard, 1981). Some authors treat the behavioral estimates in a number-to-position task as a direct and linear measure of the underlying numerical psychological scale. In other words, they assume the existence of a direct mapping from the number representation to response locations. They further advocate that number-to-line findings add to the list of evidence for an inherent relation between human's numerical and spatial representations. Specifically, they theorize that numbers are represented on a spatially oriented "mental number line" (Dehaene, Bossini, & Giraux, 1993; Zorzi, Priftis, & Umiltà, 2002; Hubbard, Piazza, Pinel, & Dehaene, 2005; Opfer & Siegler, 2007; Siegler. Thompson, & Opfer, 2009; Dehaene et al., 2008, 2009; de Hevia & Spelke, 2009, 2010; though, for a rebuttal, see Cantlon, Cordes, Libertus, & Brannon, 2009; Núñez, 2011; Núñez, Doan, & Nikoulina, 2011; Núñez, Cooperrider, & Wassmann, 2012).

The authors advocating the mental number line propose that such representation is initially an inaccurate, approximate estimation system, whose output is illustrated by the young children's logarithmic pattern of estimates in number-to-line tasks. A logarithmic representation with constant variability would be advantageous in cases where numerosities are unfamiliar, because it allows a better discrimination between lower values (Opfer & Siegler, 2007). The transition between the logarithmic and the linear representation of numerical magnitude supposedly occurs by experiencing information that does not match the earlier logarithmic representation (Opfer & Siegler, 2007). Researchers have pointed out that schooling, or formal education, plays an important role in this representational shift. They highlight the importance of learning a verbal counting routine and familiarity with numbers (Lipton & Spelke, 2005; Le Corre & Carey, 2007; Ebersbach, Luwel, Frick, Onghena, & Verschaffel, 2008) and of an extensive training with visual symbols, rulers, and left-to-right oriented physical number lines, as well as other measurement devices (Berteletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010; Núñez, Cooperrider, & Wassmann, 2012). Once the child has learned a linear representation regarding smaller numbers (e.g., the integers between 1 and 10), he/she may extend this representation towards a larger range numerical context, which in the meanwhile has become incompatible with a previous logarithmic representation.

Results from SNARC, Line-Bisection as well as Number-to-Position procedures have been advanced as the basis for the conception that number representation is spatially organized. However, the SNARC effect, studied mostly in educated human adults, seems to be largely affected by instructions and other procedural manipulations. As such, an alternative view to the mental number line hypothesis has defended that the SNARC effect "reflects recent spatial experiences, cross-modal associations, and long-standing directional habits but not an attribute of the number concepts themselves" (Gevers & Lammertyn, 2005; Fischer, 2006; Fischer, Shaki, & Criuse, 2009; Santens & Gevers, 2008).

Likewise, the interpretation of Line Bisection results as evidence for an unlearned and privileged association between number and space has been criticized.

Advocates of the mental number line would interpret the bias in the following way: in a logarithmic representation of numbers, larger numbers are compressed to the right and the midpoint numerosity is closer to the largest number than to the smallest number (see, e.g., the logarithmic line at the right bottom of Figure 4). As a result, the subject perceives the line's subjective midpoint towards the largest number (Figure 2) (de Hevia & Spelke, 2009; Bulf, Cassia, & de Hevia, 2014). However, as it has been pointed out by Núñez, Cooperrider, and Wassman (2012; see also Gebuis & Gevers, 2011), because in both SNARC and Line Bisection procedures the participant emits a single spatial response, it goes against the own definition of mapping ("mapping", 2006) to describe the reported number-space associations as evidence of a "spontaneous mapping of number onto an oriented space" (Bulf, Cassia, & de Hevia, 2014). We must then turn to the Number-to-Position procedure, which seemingly conforms to the definition of mapping, given that numerosities should be assigned to particular positions along the line (Núñez, Cooperrider, & Wassman, 2012; Siegler & Opfer, 2003).

6. Dehaene's Number-to-Position study with the Mundurucu

The strongest evidence for a universal innate logarithmic mental number line comes from Dehaene and colleagues (2008)' study with the Mundurucu, an Amazonian indigenous tribe with a reduced lexicon of number words and little or no experience with rulers, maps, graphs and other measurement devices (Dehaene, Izard, Spelke & Pica, 2008). Dehaene and colleagues adapted Siegler and Opfer's (2003) number-to-line task to test the representational systems of Mundurucu adults and children, contrasting their performance with North American adult participants. The novelty of their study was that they tested not only symbolic but also nonsymbolic numerical stimuli. Specifically, the numerosities within the range of 1 to 10 were presented in two symbolic conditions: spoken Mundurucu (Numerals[1,10]) and Portuguese words for Mundurucu participants (American participants heard number words in English and Spanish). Numerosities in the 1 to 10 range were also presented in two nonsymbolic conditions: arrays of dots (Dots[1,10]) and sequences of tones (Tones[1,10]). There was also a nonsymbolic larger range condition, with visual arrays ranging from 10 to 100 dots (Dots[10,100]).

Participants were presented with a line segment in a computer screen flanked at its endpoints by 1 and 10 dots or 10 and 100 dots, depending on the numerical range being tested (Figure 5).

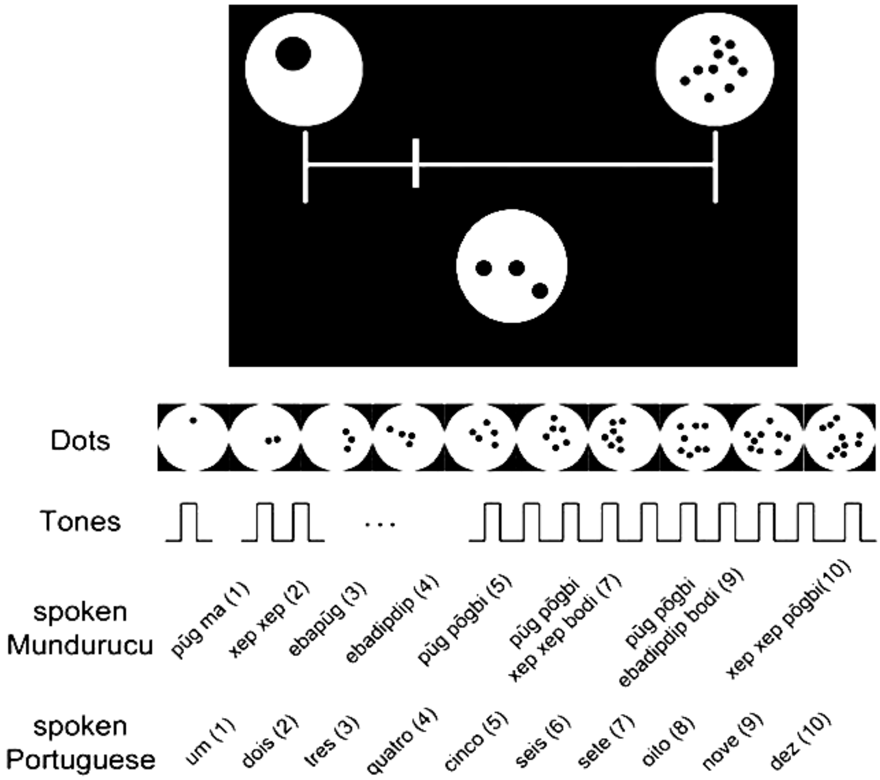


Figure 5. Dehaene and colleagues (2008)' Number-to-Position procedure. The upper portion of the figure represents one trial in the Dots 1-to-10 condition. Participants saw a line, flanked by one dot at its leftmost position and by ten dots at its rightmost position. A numerical sample was presented centered below the line. Participants were also tested with sequences of tones, spoken Mundurucu and spoken Portuguese words.

At the beginning of the session, the participant was told that the line was a path that went from the smallest to the largest number. Next, the participant completed two training trials to learn the anchor-endpoint mappings (e.g., "1-leftmost", "10-rightmost" during the Dots [1,10] condition). Afterwards, the testing phase started during which both the anchors and novel, intermediate numerosities were presented. The participant had to point with a pencil at a location along the line displayed on the computer screen. The experimenter would then use the mouse device to click at the pointed location, so that a cursor vertically bisecting the line was depicted at the selected location. Having confirmed the placement of the response, the experimenter would press a key to begin a new trial. American participants, however, responded entirely by themselves.

The study aimed to investigate whether Mundurucu, whose culture had few counting and measurement tools, would spontaneously map numerosities onto space and, in the affirmative, to investigate the structure of the representational scale. The results are presented in Figure 6.

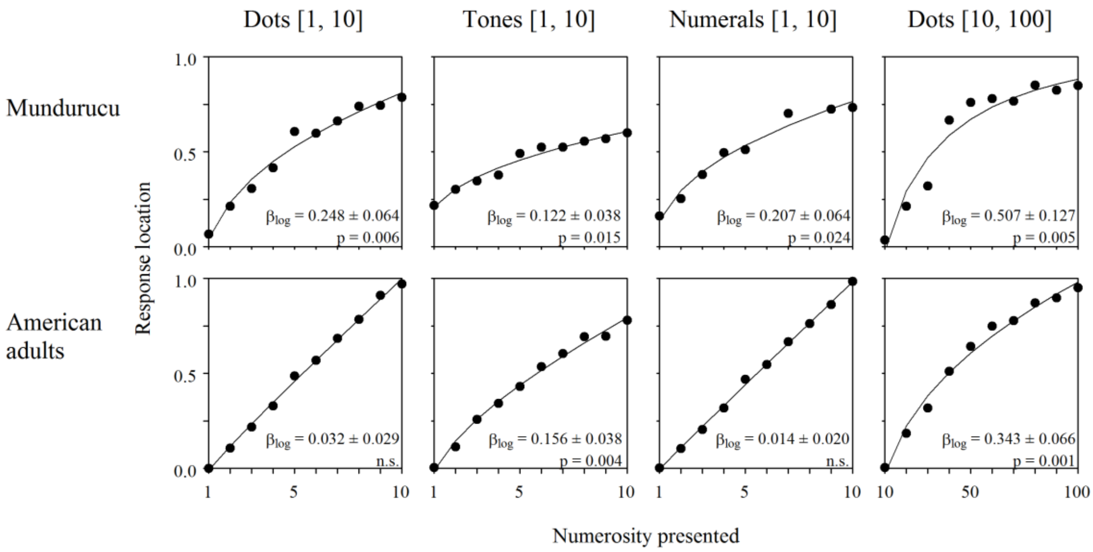


Figure 6. Results from Dehaene and colleagues (2008)' study.

Munducuru's mean response locations increased as a function of numerosity. Moreover, their group curves were negatively accelerated, and ordinary least squares (OLS) multiple regression analyses showed that a logarithmic regressor, rather than a linear one, best described their performance in all stimulus conditions, except in the case of Portuguese numerals (not depicted in Figure 6). In this case, separation of Mundurucu participants according to the education level showed that three or more years of education were associated with a trend towards a linear performance in a Portuguese numerals task, but this variable did not affect performance in the remaining stimulus conditions.

As for the American adults, their performance was best described by the linear regressor in the symbolic conditions (English and Spanish number words) and in the smaller range dots condition (1-10). A logarithmic compression was in effect when they heard 1-10 tones or saw 10-100 dots, arguably much harder to count stimulus conditions. This was addressed as evidence that, as previously suggested in studies with symbolic conditions, linear representations are available only when numerosities are presented symbolically or, in nonsymbolic cases, when their magnitude can be precisely assessed.

In conclusion, Dehaene and colleagues' results with a population lacking western education and cultural practices (e.g., use of rulers, counting routines) , as well as a specific lexicon for numbers, were taken as evidence for an inherent association between numbers and space, which emerges ontogenetically and may be later affected by culture (de Hevia, Girelli, & Cassia, 2012).

7. Current thesis

According to the mental number line premise, immersion with measurement tools and counting routines is responsible for the linearization of a prior logarithmic-like pattern of responses. A resulting conjecture is then that young children's mapping of nonverbal (nonsymbolic) numerosities must differ from adults'. In a somewhat simplistic formulation, the older the child is, the more he/she has acquired experience with rulers, graphical depictions of numerical quantities, and the higher his/her proficiency in counting and verbal estimation abilities. However, as in Siegler and Opfer (2003)'s seminal study, number-to-position studies with children have tested them with spoken number words or printed Arabic digits.

The experiments we present in the current dissertation initially started because we wished to replicate with preschool children Dehaene and colleagues (2008)'s number-to-position experiment with both symbolic and nonsymbolic numerical samples. Specifically, we wanted to contrast preschoolers' performance with that of adults in the same stimulus conditions as in Dehaene and colleagues (2008)'s experiment with the Mundurucu.

In Study 1, our question was whether we would obtain evidence consistent with a developmental log-to-linear shift while implementing some experimental controls such as recurring to a computerized experimental procedure to decrease the possibility of an effect of the interaction between participants and the experimenter (particularly in the case of preschoolers). Thus, in all our experiments, we tried to decrease as much as possible the experimenter's involvement during the trial events between the presentation of the numerical sample and the participant's response.

The major contribution of Study 1 was the incongruence between different levels of analysis when inspecting preschoolers' number-to-position performance. Inspection of single-trial responses was at odds with the mean group function. The importance of this finding is increased because a description of individual performance has been absent in prior verbal (symbolic) number-to-position studies.

It was the results from Study 1 that determined the direction followed in the remaining studies. Because of preschoolers' failure in the task's mapping criterion, we empirically addressed which prior learning may enhance responding along the extent of the spatial response *continuum*. As such, the main goal in Studies 2, 3, and 4 was to investigate how a response continuum comes under the control of the numerical stimulus continuum. In these three studies, we investigated the effects of three pre-training histories on the Number-to-Position performance of preschoolers and young adults. In Study 2, prior to a Number-to-Position testing, we ensured that participants had learned to respond along the bar as a function of increasingly darker stimuli. We aimed to implement such Brightness-to-position pre-training similarly to how behavioral psychologists teach "continuous repertoires" to animals and young children (Wildemann & Holland, 1972). This way, both in the Brightness pre-training and the ensuing testing with Number-to-Position tasks, the values in the stimulus and response dimensions could be ordered along a continuum of increasing magnitude. In Study 3, the pre-training isolated the "mechanical" component of serial responding along the response bar (response dimension). Namely, participants were pre-trained to respond along the bar as a function of non-ordered images, and afterwards entered the Number-to-position testing. With a focus on the numerical stimulus dimension, in Study 4 we investigated if increasing sensitivity on numerical discrimination along the tested numerical interval would affect Number-to-Position performance.

A pervasive concern across the four studies was evaluating how well average data represented individual performance. We believe that a major contribution of the current thesis is the challenge it poses to the theoretical proposal of a privileged number-space association. The other contribution is undoubtedly the invitation for a focus on individual performance, extended even to previously published Arabic digits Number-line studies, whose main focus of analysis was the average group curve.

CHAPTER II : STUDY 1

Preschoolers' categorical vs. adults' continuous mapping patterns in symbolic and nonsymbolic number-to-position tasks

1. Introduction

The present study aims to contrast human adults and preschool children's performance in number-to-position tasks. The linearization of an initial logarithmic mapping across development seems to be an established finding. However, except for Ebersbach and colleagues' (2008) study, all number-to-line procedures with western preschoolers have tested solely symbolic conditions, i.e., Arabic digits and/or spoken numerals. In Ebersbach and colleagues' procedure, children from kindergarten to second grade level were presented a horizontal 80 cm wooden stick, along which a pointer could be moved. Two cards with both printed Arabic numerals and an equivalent number of dots were placed at the endpoints of the stick. The stick's endpoint at the children's left side had the digit "1" and one dot, and the stick's endpoint at their right side had the "100" digit and one hundred dots. The task was presented as a story in which moving the pointer along the bar indicated the number of chocolate pieces required for a certain number of guests at a birthday party. After training the anchor mappings, children were tested with other numerosities and no feedback was given to their responses. During the test, the experimenter showed cards with a certain number of dots and a printed digit which he read out loud. The authors found that their model of two linear segments provided the best overall description of children's performance, but between the linear and logarithmic functions, the logarithmic function fitted the data significantly better in the two youngest school age groups, and the linear in the oldest ones.

Nonetheless, the joint presentation of verbal and nonverbal information does not allow for an identification of which features of the sample stimulus controlled the responses. When children see the set of dots, read the numeral, and are told the number by the experimenter, it is not clear if they are being guided to respond based on their own cardinal meaning of a number word (Wynn, 1992), or if they are attending to the non-verbal magnitude? Considering other numerical task findings with young children (de Hevia & Spelke, 2009; Holloway & Ansari, 2009), an effect of the numerical sample being presented solely as a numeral or presented nonverbally may be expected. Thus, in the current number-to-line study, similar to Dehaene and colleagues' study with the Mundurucu, <u>we tested children and adults in separate symbolic (i.e., numerals) and nonsymbolic stimulus conditions</u>.

Several standing issues in the current literature are worthy of attention. For example, aside from testing with symbolic numerals, many procedures are highly dependent on interactions between the experimenter and the participant. The experimenter usually reads each numerosity to the participant and, possibly because the numerals and the response line are printed on paper, remains at the side of the child during the entire session. Furthermore, the experimenter prompts during response moments, is responsible for progressing across trials and, at times, even records the responses (Siegler & Opfer, 2003; Siegler & Booth, 2004; Opfer & Siegler, 2007; Ebersbach et al., 2008; Berteletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010; Barth & Paladino, 2011; Dehaene et al., 2008). Consequently, to minimize experimenter bias, we implemented a fully computerized procedure.

Another, even more compelling issue, concerns data analysis. In previous number-to-line studies, the median or mean from different age groups, and/or cultures, are either the sole or the main unit of analysis. Individual analysis in number-to-line tasks, when conducted, usually consist of simple regression analyses using linear ($Y = m*X+b$) and logarithmic ($Y = m*Log_{10}(X)+b$) functions. Similar to Siegler and Opfer (2003)'s study, each participant is then classified as either "No representation" -- if both regressions failed to reach significance -- or as "Linear" or Logarithmic", depending on whether the statistically significant best fitting model was the linear or the logarithmic. The most striking feature is the considerable number of preschoolers whose performance is neither fitted by a linear nor a logarithmic function in studies with Arabic numerals. For example, in Berteletti and colleagues' study, children from 4 to 6 years of age were tasked with line-mapping Arabic numerals from 1 to 10 and from 0 to 100 (Berteletti, Dehaene, Piazza, & Zorzi, 2010). While the well-established pattern of a log-to-linear shift as a function of age was replicated, about 24% of the children in the Numerals[1,10] and 44% in the Numerals[0,100] condition were neither characterized by a linear nor a logarithmic function. It seems reasonable to affirm that such level of analysis may have disregarded interesting individual patterns, which would arguably tell us more about children's understanding of number and space.

Another concern is that, though participants may fail to respond along the response bar, inspecting only average curves could suggest otherwise. That a continuous repertoire might not occur at an individual level, even if an average curve would suggest otherwise, is a warning to take into account in the stimulus control research (Estes, 1956; Migler, 1964; Bickel & Etzel, 1985). To characterize the performance of "No representation" cases, and verify whether average location depicts single trial responses, one must conduct at least an initial inspection of individual scatterplots.

How uncommon are non-*continuum* patterns? Did they occur in previous studies and if so, how were they dealt with? The above mentioned study of Berteletti and colleagues provides no further clarification except for a footnote stating that "among the children classified as not having a linear or a logarithmic representation, *some* use evident nonnumerical strategies such as alternating between the left and right marks on the lines" (Berteletti, Dehaene, Piazza, & Zorzi, 2010; italics ours). Barth and Paladino's (2011) study, also with Arabic digits between 1 and 100, was the first to present individual scatterplots, with both single trial and median location of responses. In their preschooler group, 5 out of 21 five-year olds "placed at least half of their estimates at the midpoint of the number line and ⁄or produced estimates unrelated to the presented numeral", despite having been previously trained on the anchors and middlepoint mappings. Still, since the experiment aimed to compare the authors' power function model (Barth & Paladino, 2011; Barth, Slusser, Cohen, & Paladino, 2011; Hollands & Dyre, 2000) with the log-to-linear shift model (Opfer, Siegler, & Young, 2011; Siegler & Opfer, 2003), they only verified how the inclusion of these non-*continuum* cases affected the curve fitting results. Nonetheless, their individual scatterplots raise the question of whether non-continuum cases occurred in other verbal number-to-line experiments and, in the affirmative, how were they handled when carrying out group analyses.

In the "Supplementary Online Material" accompanying Dehaene and colleagues (2008b)' study, the authors mention that about 37% of the Mundurucu participants emitted a "*bimodal*" pattern of response. The "*bimodal*" designation was employed to name those participants who responded less than 20% of the time at the middle region of the response bar. In other words, a considerable percentage of the Mundurucu failed to use the full extent of the spatial *continuum*, restricting their responses to the bar's endpoints. Regrettably, the authors did not distinguish between stimulus conditions nor did they further investigate this result. Recall, however, that the Mundurucu tribe findings have been interpreted as supporting an inherent, universal number-space mapping (de Hevia, Girelli, & Cassia, 2012). This claim has been challenged both by a re-analysis of their data and by a study with schooled and unschooled adults from the Yupno tribe, in Papua New Guinea (Núñez, 2011; Núñez, Cooperrider, & Wassman, 2012). In Núñez and colleagues (2012)' study, even though the group median of response locations showed a log-like pattern, unschooled Yupno adults did not respond to locations other than the anchor endpoints when tested in the Dots[1,10], Tones[1,10], and Numerals[1,10] conditions. In contrast, schooled Yupno and Californian control participants employed the full extent of the response line.

Returning to Dehaene and colleagues (2008)' Supplementary Material, when the data from the Mundurucu adults is analyzed separately for the different schooling groups (Figure 1), one observes that responses of the critical subgroup, the unschooled Mundurucu, are not differentiated for numbers 1, 2, and 3 in the Tones[1,10] and spoken Mundurucu words (Numerals[1,10]) stimulus conditions.

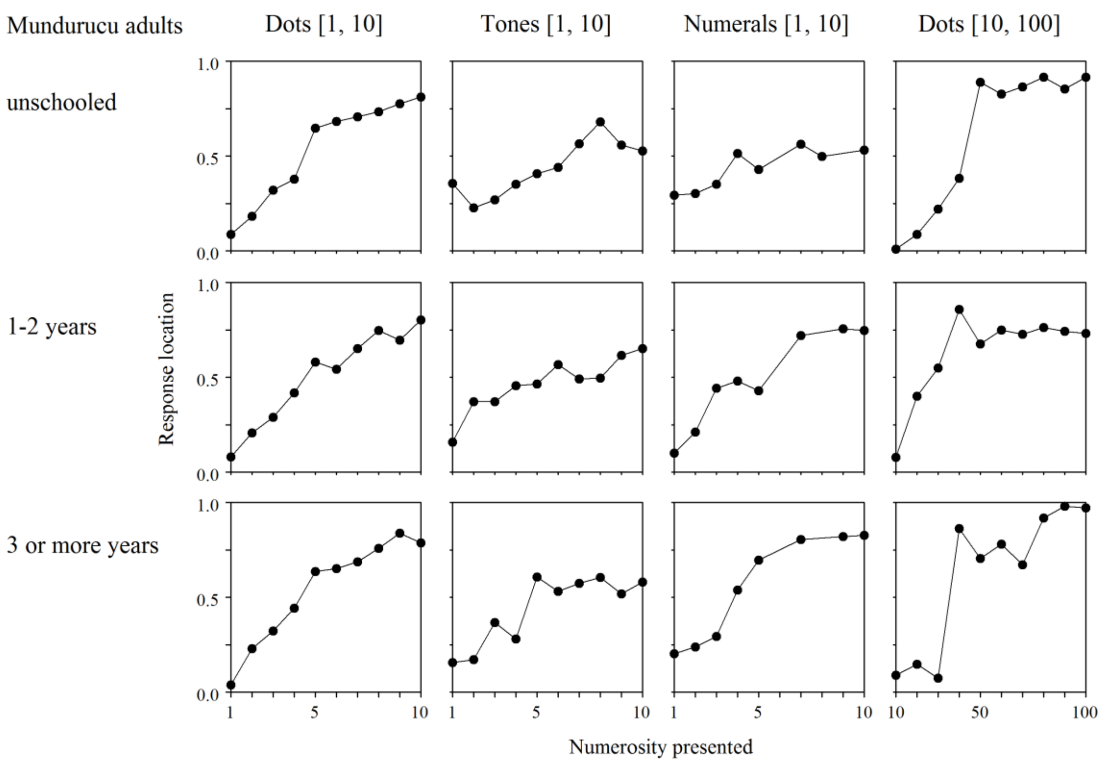


Figure 1. Average response locations of Mundurucu adults separated by schooling sub-groups: unschooled, with 1 to 2 years of formal schooling, or with 3 or more years of schooling. Note how flat the curves are in the unschooled group, particularly in the Tones and Mundurucu number words conditions.

In the Tones[1,10] condition, responses to number "1", the smallest anchor, are actually further along the line than to numbers 2, 3, and 4. If the numbers were represented logarithmically, we would expect these smaller values to have been the easiest to discriminate (Siegler & Opfer, 2003). In addition, many sub-groups' response curves are practically flat and suggest that throughout the tested numerical range, subjects continued to respond at the line's middle region.

Because they report group average data, the logarithmic-like patterns in previous number-to-line studies might, to an unknown extent, have resulted from the participants distributing their responses amongst the trained anchor-positions. And although "logarithmic responding may occur whether or not the participants' responses were distributed unimodally or bimodally on the response continuum" (Dehaene et al., 2008, OSM), in the latter case the dependent variable -- mean location -- should not be considered as the characterization of an actual location of responses along the bar, with the resulting log-like feature of average location being an averaging artifact. Thus, a final aim of our study is to analyze individual performance. In the following study, we intended to address how well the measure of central tendency reflects individual patterns, as well as the main differences across age groups and stimulus modalities.

2. Method

2.1 Participants

Eighteen Portuguese pre-schoolers (8 girls and 10 boys) and eighteen Portuguese young adults (8 women and 10 men) participated in the current study. The mean age of preschool children was 5.64 years (SD = 1.32) and the mean age of adults was 21.72 years (SD = 2.01). Informed consent was given by all adult participants and the children's parents.

2.2 Procedure

Participants were seated in front of a laptop (screen size 13.3") in a quiet room, a separate room in a kindergarten in the case of children, and in a community center in the case of adult participants. Participants used a computer mouse when controlling the movement of the cursor on all x/y screen positions. Prior to the experiment, children participants interacted with the experimenter during their usual school activities and played with the computer in a drawing program (MS Paint), which required them to use the mouse.

An experimental program written in Visual Basic controlled all session events and recorded participants' responses. The experimenter remained in the room, seated about 1.5 m behind the participant to remain out of sight and prevent response bias. Each participant completed five experimental sessions, one on the *Brightness* Testing and four on the *Numerosity* Testing for four stimulus modalities.

Brightness Phase. The first experimental phase aimed to ensure that all participants had had some exposure to a computerized task instructing them to respond along a spatial *continuum* as a function of a stimulus *continuum* (stimuli varying in brightness, a non-numerical dimension).

Training. Participants were presented with a rectangle response bar (17 cm width x 1 cm height) located 10 cm below the upmost part of the screen. The response bar background was colored with a grayscale gradient from white to black, in the left-to-right orientation of the figure. The experimenter stated that the response bar was a path that went from the "least dark" to the "most dark" color and, while saying that, moved her finger along the bar. Participants were further informed that prior to putting on the headphones and beginning the experiment, they would start by learning what stimuli belonged to the start and end positions of the path.

Each trial started with an entirely white screen. Following an Inter-Trial-Interval (ITI) of 10 seconds, a circular image of the cookie monster's head (diameter 3 cm) appeared at a random location. A mouse click on the image removed the image, turned the screen blue (RGB color (0, 78, 152)) and triggered a 16 ms auditory 'click' stimulus. Then, after a delay of 100 ms, a sample stimulus appeared centered on the screen. Sample stimuli were uniformly colored squares (side 6.35 cm), either white ("least dark") or black ("most dark"). The sample stimulus remained on the screen for 1500 ms, after which it disappeared and the response bar appeared positioned at about 10 cm below the top of the screen. The trial ended with either a mouse click at any location along the bar (the choice response), or after 10 s had elapsed, whichever happened first. An auditory "click" accompanied the choice response and the program saved choice latencies and the x/y coordinates within the response bar. If the sample had been white, the correct response was a mouse click at the leftmost portion of the bar, whereas the rightmost portion of the bar was correct following a black sample stimulus. The acceptable region was 1/10 of the bar length (i.e., 1.7 cm).

A correct response turned the screen white and presented a yellow star image (square side 7.94 cm), announcing the recorded verbal feedback "Very well!". With preschool participants, the experimenter delivered a small portion of a marshmallow in a cup next to the participant. Time of delivery of the treat was signaled to the experimenter in another screen that was concealed from the child. A correction procedure was in effect following incorrect responses and it consisted of repeating the trial with the wrong portion of the bar occluded with a dark blue rectangle during the choice moment. This step ensured that only the correct area of the response bar could process mouse click events during the correction trial.

Training progressed until the participant completed at least twenty trials (not counting the repeated trials) and had reached the criterion of four consecutive correct responses for each anchor stimulus. Adult participants could end their training after completing ten trials with four consecutive correct responses for each anchor stimulus.

Testing. The experimenter told the participants that, in addition to the less dark and the darkest stimuli, they would also see values of brightness ranging from one to the other. Their task was to decide where to place the sample along the response bar, knowing that stimuli got "darker" as one moved from the left to the right side of the response bar (the experimenter moved her finger along the bar while she spoke).

The trial events were similar to those in training, with the following changes: (i) the sample stimulus (i.e., the background color of the square) could have been one of six brightness values, namely the previous two anchors and four grays; (ii) when the stimulus was one of the two anchors, correct responses were followed by the previously described feedback events with a maximum probability of .67, depending on accuracy, with no correction procedure in effect; (iii) responses following gray samples, given that they did not occur at the anchor sites, were non-differentially reinforced (feedback plus treat in the case of children, only feedback in the case of adult participants) with a probability of .33. Testing ended after 42 trials, 12 exemplars of each anchor, 6 exemplars of gray1 (RGB color (178, 178, 178)), 3 exemplars of gray2 (RGB color (127, 127, 127)) and of gray3 (RGB color (102, 102, 102)), and 6 exemplars of gray4 (RGB color (51, 51, 51)). The stimuli were presented in random order.

***Numerosity* Testing Phase**. Children began the Numerosity Testing Phase after a minimum interval of one day. Adult participants, however, were free to continue immediately after the Brightness phase. The samples were numerical stimuli presented either as sets of dots ranging from 1 to 10 dots (Dots[1,10]), sets of dots ranging from 10 to 100 (Dots[10,100]), a sequence of 1 to 10 tones (Tones[1,10]), and verbal Portuguese numerals (Numerals[1,10]). Each sample condition was tested in a different session. Order of sessions was counterbalanced for the adult participants. However, due to a computer programming error preschoolers started with either the Dots[1,10] or the Tones[1,10] condition and the remaining conditions were counterbalanced.

In both Dots conditions, the numerical sample was a white square, with the same dimension and located at the same position as the one in the preceding brightness session. The white square contained a variable number of black dots. Two types of sets were created, one presented during the anchors training and consisting of only the two anchor numerosities, and the other presented during testing and consisting of the number of dots ranging from the smallest to the largest anchor. For the initial training of anchors, the sets were matched on total surface area so that individual dot size in the smaller anchor sets was larger than the one in the larger anchor sets. During the testing phase, individual dot size was constant for all numerosities, with the total surface area covarying with numerosity. In the Dots [1,10] session, the numerical sample was a white square that could have contained 1 to 10 black dots, each 0.7 cm in diameter. For the Dots[10,100] session, the white sample square could have contained from 10 to 100 black dots, in steps of 10, each dot with a diameter of 0.19 cm. For the Tones[1,10] session the numerical sample was a sequence of 447 Hz tones. For the smallest sample, “1”, the single tone duration was 500 ms. Tone duration decreased progressively with the number of tones in the sample, with a constant gap between tones within the same sequence, and such that the total duration remained constant across all numerosities. For the larger sample, “10”, each tone lasted 40 ms with the silent gap of 100 ms between tones. For the

Numerals[1,10] session, the numerical samples were prerecorded-Portuguese number words from one to ten.

Each numerical session started with the training of the correct location of anchor stimuli, namely, the smallest and the largest numbers. The experimenter would state that they would start a new game with numbers. The sample would be a numerical stimulus (number of dots, of tones, or numerals). The response bar was a path that went from the less numerous to the most numerous location and, while stating that, would move her finger along the figure. She then said they would begin by learning about the numbers belonging to the start and end positions of the path, then they would put on the headphones and start the experiment. All session events, such as the ITI, the subject's response to start the trial, the spatial and temporal configurations of sample and choice stimuli presentation, and the reinforcement contingencies, were similar to the ones described above for the Brightness Phase. There were a few adjustments such as that during the sample a numerical stimulus was presented, and the response bar displayed a uniform white background. Following anchor training, participants were told they would see many numbers, ranging from the minimum to the maximum, and would have to decide where to place them along the path (the experimenter moved her finger along the response bar, in the left-to-right orientation). The testing phase was composed of 42 trials, with five presentations of each numerical anchor and four presentations of the eight arithmetically spaced intermediate numerosities. During the testing phase, the reinforcement and feedback contingencies remained the same as in the Brightness testing phase.

3. Results and discussion

3.1 Brightness session

Response locations were measured as the relative distance from the response bar left endpoint, so that they could range from 0, when the mouse click occurred at the leftmost location, to 1, when the mouse click occurred at the rightmost location; a response location of 0.5 corresponded to a click at the midpoint of the response bar.

As depicted in Figure 2, during the testing phase subjects responded along the bar, progressively to the right, as the sample got darker (significant Pearson's correlation coefficients between sample and group mean response location in the Children, $r(4) = 0.98$, $p = .001$, and Adults, $r(4) = 1.00$, $p < .001$, groups).

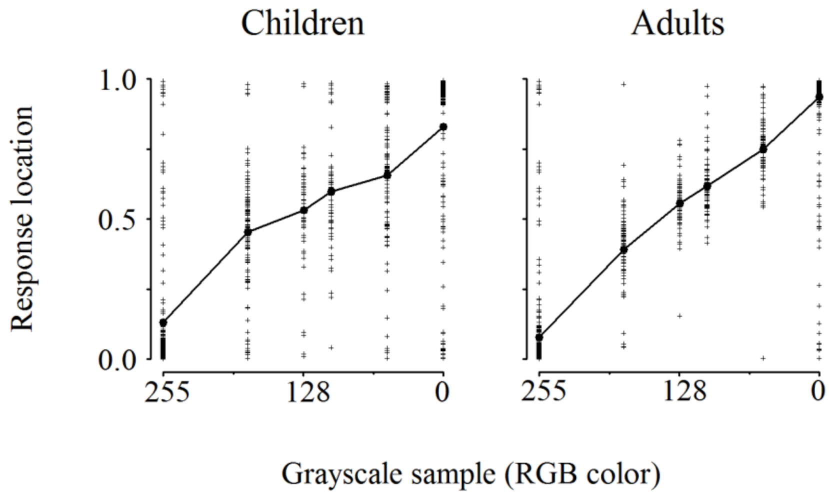


Figure 2. Children and adult participants' response locations as a function of the sample's brightness (ranging from white (RGB 255, 255, 255) to black (RGB 0, 0, 0)). Each small cross point represents a single trial response and the dot-connected line the mean location from all trials taken together.

Children made more errors in the trained anchor mappings (7 children failed the "white", and 12 the "black" sample) than the adults (7 subjects in the "white", 9 in the "black" sample). When tested with gray stimuli, both children and the adults selected locations other than the trained anchors. This is confirmed by the results of a mixed between-within ANOVA with a Greenhouse-Geisser correction, with age group as the between factor (2 levels) and of grayscale sample (6 values) as the within factor. Not surprisingly, there was a main effect of sample value in the response location, in that the darker the stimuli the higher the selected locations ($F(1,4.33) = 285.209$, $p < .001$). Also, a significant main effect of age ($F = 5.489$, $p = .021$), showed that adults selected more locations further along the bar than children did. This result is interpreted when we also consider the significant interaction effect ($F = 6.529$, $p < .001$). As illustrated in Figure 2 by the larger spread of response locations, children tended to concentrate their responses at the midpoint of the bar and the anchor positions more than adults. As for adults, their curve approached the figure's diagonal more than children's did and their response locations were broader with more of a continuous use of the response bar,

3.2 Numerosity sessions

3.2.1 Group analyses

In both age groups and for all stimulus modalities, mean response location increased with the numerical sample (significant positive Pearson's correlation coefficients between stimulus numerosity and mean response location; $r(8) > 0.91$, $p < .001$). A main interest of this study was to identify how group results relate to that of the existing literature. Therefore, OLS multiple regression models contrasted the contribution of a non-linear component over and above a linear regressor to judge whether a logarithmic model, rather than a linear model, would prove to be a better descriptor of the spatial mapping.

Figure 3 presents for each condition, mean response location (± 1 SEM) as a function of presented numerosity. We also plotted the results from Dehaene et al. (2008)'s study contrasting Mundurucu adults and children with American adults, as well as American adults' average locations, from Núñez, Doan, & Nikoulina (2011)'s study (refer to Appendix A for a table with the complete results from ours, Dehaene et al.'s and Núñez et al.'s multiple regression analyses. The Appendices for each study are presented at the end of this Dissertation).

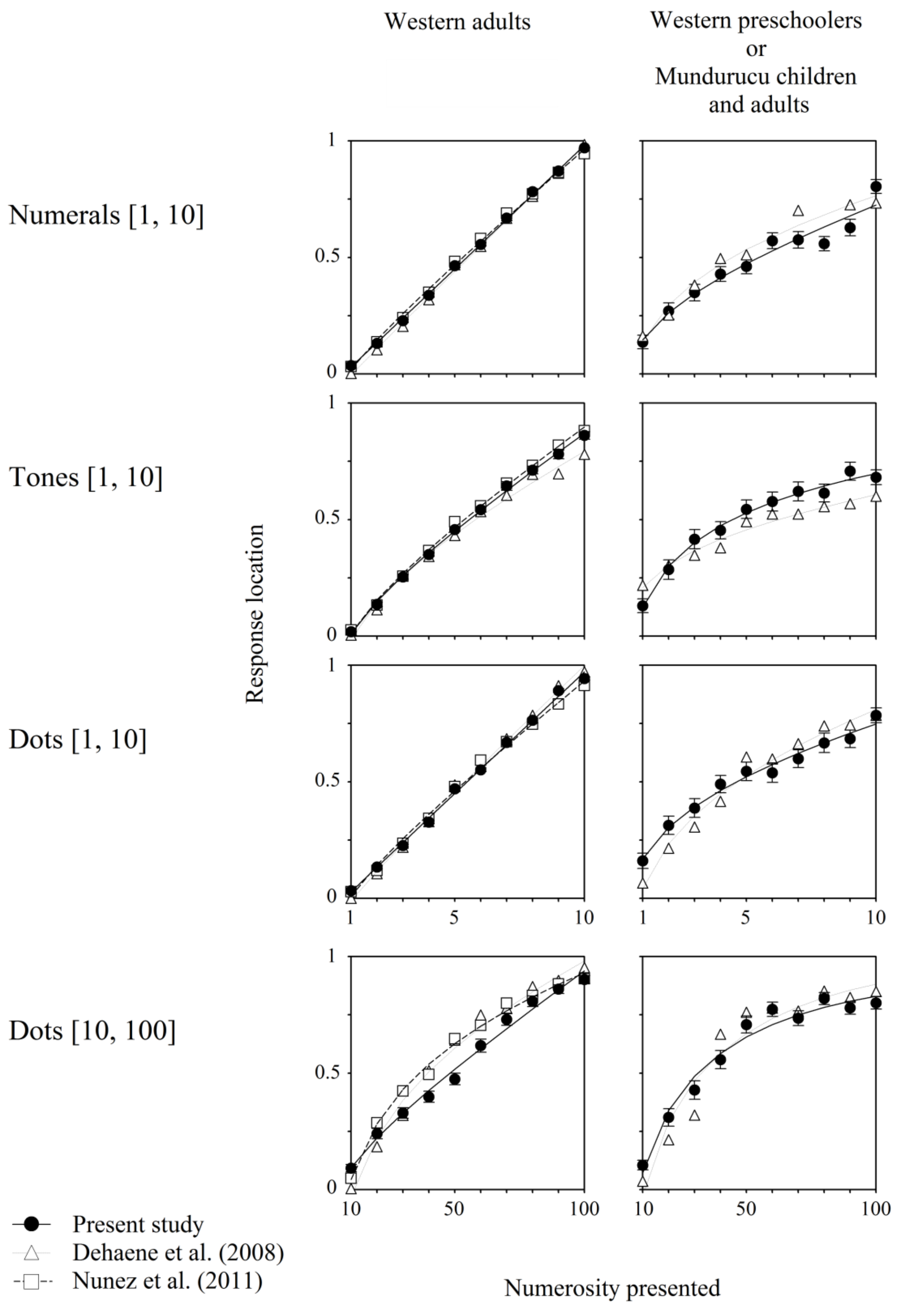


Figure 3. Numerical mapping of Western and Mundurucu children and adults. Each datapoint is the mean location of all participants' responses taken together, with bars representing the SEM. Data from the current study is depicted by black circles: Portuguese adults at the left column, prescholers at the right column. Data from Dehaene et al. (2008) is depicted by triangle data points; left: American adults, right: Mundurucu children and adults. Square datapoints represent American adults' performance in Núñez et al. (2011)'s study. Lines are the multiple regression fitted functions (see Appendix A).

<u>Adults</u>. Adults' performance is depicted by the filled circles displayed in the left column of panels in Figure 3. As indicated by nearly a straight line in the Numerals[1,10] condition graph, when adults were presented the spoken number words from 1 to 10 (i.e., verbal/symbolic numerosities), they placed them subsequently along the bar with approximate equidistant spacing between them. However, when the same quantities were presented as sequences of tones (i.e., non-symbolic), the series became more curved.

This visual inspection is backed by the results of the multiple regression, which confirms that the contribution of the logarithmic regressor was significant for the Tones[1,10] condition (β_{log}: p = .001; β_{lin}: p < .001), and non-significant when the numerical stimuli were number words (Numerals[1,10]: β_{log}, n.s.; β_{lin}: p < .001). The graphs also depict the mean data from Dehaene and colleagues (2008) and Núñez and colleagues (2011)' studies with western adults, as well as their fitted multiple regression curves. Similar patterns are observed among the different studies. Thus, our results replicate their findings in that number mapping of number words is solely linear, whereas when participants are presented with the same numerosities in the form of sequences of tones, the response curves have a logarithmic component.

In the third graph, depicting adults' responses in another non-symbolic condition consisting of sets of 1 to 10 dots, the mapping is linear, both in ours and in the previous studies (β_{lin}: p < .001; β_{log}: n.s.). We thus replicated the finding that adults' performance in number-to-line tasks differs when nonverbal numerical samples are presented simultaneously (Dots[1,10] condition) or sequentially (Tones[1,10] condition).

Adults' extensive training with counting routines and accordingly, with the algorithm that consecutive integers are separated by a constant interval, has been suggested as a possible reason for the linear mapping of small numbers (Lipton & Spelke, 2005; Ebersbach, Luwel, Frick, Onghena, &Verschaffel, 2008). This explanation applies to visual 1-10 nonverbal numerosities such as sets of dots, but not to sequences of tones – which reveal a logarithmic mapping.

It is still unclear which properties of the numerical samples may be the reason for this difference. Is it a question of modality (visual vs. auditory) or mode (simultaneous vs. sequential)? Findings from other numerical discrimination procedures testing nonverbal numerical stimuli suggest that participants respond similarly to sequences of figures or tones, but differently to dots presented simultaneously or sequentially (Piazza, Mechelli, Price, & Butterworth, 2006; Nieder, Diester, & Tudusciuc, 2006; Cohen Kadosh & Walsh, 2009; Tokita & Ishiguchi, 2012; Droit-Volet, Clement, & Fayol, 2008). If the difference between the simultaneous Dots[1,10] and sequential Tones[1,10] conditions in number-to-line experiments lies in the temporal integration of the elements that compose a numerical set (i.e., mode), then sequential sets of tones and a same number of light flashes should be similarly positioned along the bar -- an hypothesis that is yet to be tested.

In the fourth graph, the Dots[10,100] condition, we observed that our participants performed differently from what was observed in the other two studies. Namely, the linear regressor was significant but the logarithmic one was not (β_{log}: n.s.; β_{lin}: $p < .001$), whereas in previous studies the opposite was found. Indeed, our initial prediction would have been for a range effect between the Dots[1,10] and the Dots[10,100] conditions, giving the well-established finding from many numerical discrimination experiments that stimulus discriminability decreases with magnitude. However, the linear patterns found both in the Dots[1,10] and the Dots[10,100] conditions (β_{log}: n.s.; β_{lin}: $p < .001$), are inconsistent with this expected effect of numerical magnitude.

Adults' recurrent exposure to horizontal small number lines during schooling and everyday experience, together with an inability to count larger sets, have been pointed out as reasons to why their mapping is usually linear in a Dots[1,10] condition but logarithmic in a Dots[1,100] condition (Núñez, Doan, & Nikoulina, 2011; Dehaene et al., 2008). Results from the single task condition in a study by Anobile, Cicchini, and Burr (2012) also challenge this account given that, similar to our study, they found that adults' average location was markedly linear, both in smaller [1,10] and larger [1, 100] dots conditions. How procedural features could account for the differences among the four studies that have tested adults with nonverbal numerosities is yet to be understood. First, they don't seem to be due to non-numerical attributes co-varying with numerosity. In our study, training stimuli were matched in total area and varied in dot size, while during testing dot size was held constant across sets, so total area co-varied with numerosity, which could have provided an additional clue about the set's magnitude. However, stimuli in Núñez et al. (2011) and Dehaene et al. (2008) studies also had summed area co-varying with number, and they found a logarithmic pattern in larger sets. Differently from the latter as well as our study, Anobile et al. (2012) implemented controls for spatial cues alternately: across successive testing blocks either total surface area or individual dot size were controlled. Secondly, we are also not able to attribute the differences to features such as the choice moment starting after the numerical sample had elapsed or occurring simultaneously with it, or even for time constraints during the sample presentation. In Dehaene et al.'s study, the sample remained on the screen until the subject had responded. In ours, as well as in Núñez et al.'s, and in Anobile et al.'s studies, the response bar appears after the numerical sample elapsed, after 1500, 1000 and 240 ms time intervals, respectively.

Until further clarification, we would tentatively hypothesize that the presentation of multiples of ten numbers may have led our participants to accurately enumerate the visual sets. Should this be true, the linear pattern in the Dots[10, 100] condition would be expected, given that previous studies with Arabic digits or spoken words show that the older and more proficient at counting, the more linear the participants' mapping patterns (Lipton & Spelke, 2005; Ebersbach et al., 2008; Ashcraft & Moore, 2012). Finally, it remains to be tested which conditions enhance verbal enumeration, and to what extent it could be a moderating variable in these and other number-to-line tasks. Future studies ought to add a verbal estimation task and inspect the correlation between verbal and spatial estimates. Another proposal would be to test for the effect of an articulatory suppression task concurrently with the presentation of the numerical stimuli (e.g., Rattat & Droit-Volet, 2012).

<u>Children</u>. Children results are depicted in the right column of Figure 3. Visual inspection suggests a curvature in all lines. The multiple regression analysis shows that when the numerical stimuli were presented as Portuguese words, the spatial mapping was linear, rather than logarithmic (Numerals[1,10]: β_{log}, n.s.; β_{lin}: $p < .001$), but the logarithmic scaling effect was significant for all the other stimulus conditions (Tones[1,10]: β_{log}: $p < .001$; β_{lin}: n.s., Dots[10,100]: β_{log}: $p < .001$; β_{lin}: n.s., Dots[1,10]: β_{log}: $p < .01$; β_{lin}: $p = .05$.). In other words, children responded linearly when they heard number words for smaller numerosities but responded logarithmically when presented with nonverbal numerosities, in both smaller and larger numerical ranges.

In the only study that tested western children with nonverbal numerosities in a number-to-line task, Ebersbach and colleagues found that kindergarten children (mean age = 5.25 years) preferentially mapped sets of Dots[1,100] in a logarithmic rather than in a linear pattern, even though the nonverbal quantities were presented in a compound with verbal information (a printed Arabic digit that was also read by the experimenter) (Ebersbach et al., 2008). Given that the logarithmic-like pattern presented by the kindergarten children had also been found in experiments solely with Arabic numerals, the authors proposed that both nonverbal and verbal large numerosities were similarly represented, so that presenting each of them separately should lead to comparable results. Our findings with the nonverbal dots conditions seem consistent with this proposal, as young children presented logarithmic patterns similar to other human subgroups that do not possess an extensive training in measurement tools such as horizontal rulers, and are less proficient in a counting routine.

In conclusion, at the group level of analysis, preschoolers seem to respond along the bar in a logarithmic pattern, whereas adults' performance is more linear. Because up until now, children had been tested solely with verbal numbers (Arabic digits, mostly) or in a compound with spoken number words (Ebersbach et al., 2008), our results would seem to support the notion that educational experience linearizes an innate logarithmic mental number line, as it has been defended ever since Siegler and Opfer's seminal article (Siegler and Opfer, 2003; Booth & Siegler, 2006; Thompson & Opfer, 2008; Dehaene, Izard, Spelke, & Pica, 2008; Dehaene, 2009).

3.2.2. Individual analyses

All participants learned the mapping between the smallest numerosity and the leftmost portion of the bar and between the largest numerosity and the rightmost portion of the bar. During the testing phase, 16 out of 18 children failed to preserve the correct mean location for at least one of the two anchors in the Tones[1,10] condition. This number further decreased to 14 in the Dots[1,10], 13 in the Numerals[1,10], and to 10 in the Dots[10,100] conditions (A chi-square revealed no significant differences between the conditions, $\chi^2(3) = 5.362$, $p = .147$). Adults' performance during the testing phases showed that a significant difference between conditions, with 10 adults failing the 10-anchor during the Tones[1,10] condition, 8 during the Dots[10,100], and 3 in the two remaining conditions ($\chi^2(3) = 9.5$, $p = .023$).

In accordance to the existing analyses carried out at the individual level (Siegler & Opfer, 2003), individual linear and logarithmic functions were fitted to individual average responses, and then the participants were classified as either "No representation" -- if both regressions failed to reach significance -- or as "Linear" or Logarithmic", depending on whether the significant best fitting model was the linear or the logarithmic function, respectively. According to this classification, our adult participants presented a larger percentage of linear patterns in comparison with the preschoolers (an average of 71% of adults against 32% of preschoolers), which confirms at the individual level the pattern suggested by the group level analysis. However, the most striking feature is the considerable number of preschoolers whose performance is neither fitted by a linear nor a logarithmic function. In the present study, while no adult participant failed significance at linear or logarithmic regressions, an average of 19% of preschoolers were classified as "No representation" (the conditions with a higher percentage were the Tones[1,10] and Numerals[1,10] with 28% of "No representation" cases). As mentioned in the introduction, the few studies that have carried out an individual curve fitting analyses – and where all the numerical samples were symbolic - have also reported a considerable number of "No representation" child participants.

To characterize the performance of "No representation" children, and to verify how well average location depicts single trial responses, one must inspect individual scatterplots. One concern is that, as we mentioned earlier, though participants may fail to respond along the response bar, inspecting only individual average location could suggest otherwise. The most striking feature in our data is that a large number of children participants either did not respond at portions of the bar other than the anchor positions, (bi-categorical) or responded at both anchors plus the midpoint (tri-categorical).

Consider the case of a bi-categorical mapping pattern exemplified in Figure 4 by participant C1 in the Dots[10,100] condition of our study. Average location increased as a function of numerosity. However, when we project single responses onto the y-axis, only two locations of the bar were selected: the leftmost and rightmost anchor. As such, the child kept responding at the two anchor positions in differentiated proportions. In this case, the dependent variable "average location" does not characterize an actual location of responses along the bar, and the resulting log-like feature of average location is an averaging artifact.

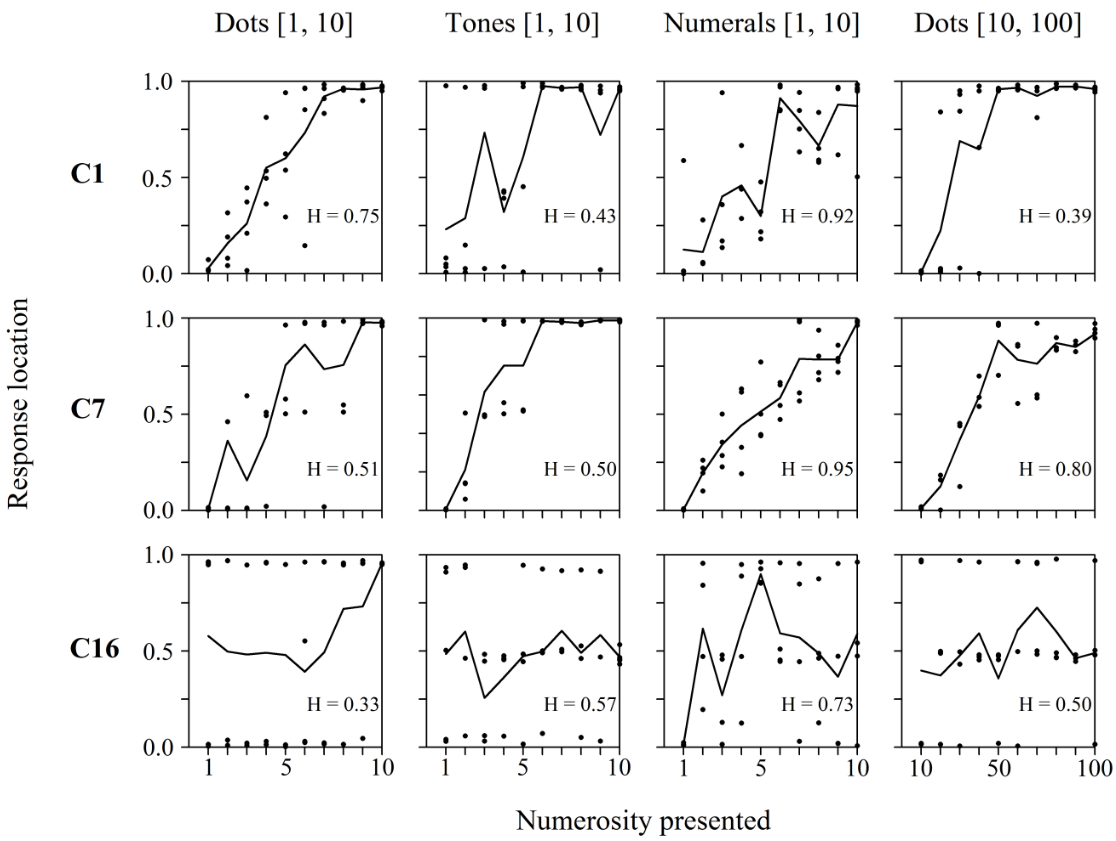


Figure 4. Results of three children, C1, C7 and C16. Each data point is the response location on a single trial; the lines are the mean locations. Each graph also presents the value of the normalized entropy of response locations (H).

Although the number-to-line task assumes a spatial *continuum* response dimension, children could be responding categorically, as in a numerical bisection task. A bisection task requires a "few/many" type of judgment and, applied to our example, the leftmost position corresponds to the "few" and the rightmost position to the "many" response options. If that is the case, then group average response locations will be, *a priori*, biased towards more similarity with a logarithmic than with a linear pattern (for psychometric curves of preschoolers tested in numerical bisections see, e.g., Jordan & Brannon, 2006; Almeida, Arantes, & Machado, 2007; Beran, Johnson-Pynn, & Ready, 2008; Droit-Volet, Clément, & Fayol, 2008). However, this possibility is more worrisome than just group results being biased towards a certain pattern because bi-categorical and tri-categorical responses go against the assumption of the number-to-line paradigm itself. That is, according to Siegler and Opfer (2003), the task demands a translation between numerical and spatial representations (i.e., mapping). Numerosities should be spatially ordered along the line, that is, they should be represented by a one-dimensional space *continuum* with a metric (Núñez, Cooperrider, & Wassman, 2012).

This survey of individual response configurations in number-to-line tasks, detected in the existing literature and presented in our data, leads us to enumerate the following patterns: responding using the extent of the response bar (continuous), responses restricted to the endpoint positions (bi-categorical), responses restricted to endpoint positions plus the midpoint of the response bar (tri-categorical). In all three cases, average response can either increase with numerosity, or be unrelated to, the numerical sample.

In the current study, however, classification is hindered due to the considerable variability between and within children participants, as illustrated in Figure 4. Some children, such as C16, failed to respond continuously throughout the experiment, with the average location of responses being unrelated to the numerical sample. Others, such as C1, alternated between responding continuously in some conditions, and bi-categorically in others (see Appendix B for all individual scatterplots). In an attempt to quantify such variability in the use of the response bar, for each participant's numerical session we computed a 10 bin histogram of all response locations. Afterwards, we estimated the Normalized Entropy ("Efficiency") of the individual histograms, according to Shannon's formula (Shannon, 1948; Shannon & Weaver; 1949): $H = -\frac{\sum_{i=1}^{10}(p_i \times log_2 p_i)}{log_2 10}$, where *pi* is the relative frequency at bin *i*. The normalized entropy measures the uniformity of the use of the response bar. It ranges from 0 -- when one segment of the bar is selected on 100% of the trials, to 1 -- when all portions of the bar are equally chosen.

Accordingly, participants who tend to restrict their responses to two or three portions of the bar presented smaller entropy scores (see Fig.3, participant C16 in condition Dots[1,10], H = 0.33). Those who responded with an almost equal frequency across the ten bins of the bar present larger entropy scores, as illustrated by participant C7 for the Numerals[1,10] condition (H = 0.95) (see Appendix C for all individual entropy scores).

A mixed between-within ANOVA with a Greenhouse-Geisser correction, with age as the between factor (2 levels) and stimulus condition as the within factor, tested for the statistical significance of age on the use of the response bar in the four stimulus conditions. Adults' normalized entropy scores (M = 0.92, SD = 0.02) are significantly larger than children's (M = 0.66, SD = 0.02) ($F(1, 2.66)=72.15$, $p < .001$). There is an overall effect of stimulus condition ($F(1, 2.66) = 6.56$, $p = .001$) and no interaction between the age group and the stimulus condition ($F(1, 2.66) = 2.11$, $p = .11$). Post hoc tests using the Bonferroni correction compared the stimulus conditions, revealing that children's entropy scores are significantly larger in the Numerals[1,10] than in the Tones[1,10] ($p = .025$) and Dots [10,100] ($p = .049$) conditions, but are not different from those in the Dots[1,10] condition ($p = .329$). No other significant pairwise comparison was found.

It must be noted, however, that the normalized entropy measure is insensitive to the degree and direction of the correlation between number and position. For this reason, entropy scores must be crosschecked with a visual inspection of the individual scatterplots. Regarding children results, the highest entropy scores in the Numerals[1,10] condition are congruent with the highest number of participants presenting a continuous pattern and, conversely, with fewer bi-categorical and tri-categorical participants. Next, by decreasing order of number of "continuous" participants, are the Dots[1,10] , the Tones[1,10] and lastly, the larger range Dots[10,100] conditions. The lowest entropy scores are associated with the bi-categorical response patterns, as the children exemplars illustrate in Fig. 3 (Appendices B and C).

The visual inspection of adults' scatterplots confirms that individual average locations result from a continuous mapping pattern. Thus, one can interpret the average statistics as representing actual locations along the response bar. As for stimulus condition comparisons, entropy scores in Dots[10, 100] are significantly smaller than in the Numerals[1,10] and Dots[1,10] condition ($p = .004$ and $p = .001$, respectively). The exception of the Tones[1,10] entropy scores being non significantly different from the Dots[10,100] condition ($p = .265$) is due to the results of participant A11 who presented a remarkably distinct pattern from the other adults. While his response locations increased continuously as a function of number, he never responded further than the response bar's midpoint (maximum response locations around 0.44), which resulted in more than half of the response bar never being selected.

4. Conclusions

Our study aimed to contrast preschoolers and adult participants' mapping performance in number-to-line tasks with symbolic and nonsymbolic numerosities. Our study was the first to test Western preschool children in a number-to-position task with nonsymbolic numerosities, in the form of sequences of tones ([1,10]) and simultaneous sets of dots in smaller ([1,10]) and larger range ([10,100]) conditions. As suggested by Dehaene and colleagues (2008)' experiment with children and adult Mundurucu participants, lacking a western education background, if linear patterns are generally associated with higher proficiency in verbal counting routines and with experience on measurement tools such as rulers, in the present study we expected adults to present a more marked linear mapping than preschoolers.

The most important feature of our study is how opposite the results are, when we compare the traditional level of analysis- the group average curve – and the individual performance. On one hand, our average results seem to agree with the existing studies documenting a developmental log-to-linear shift, and extend the phenomenon to nonverbal numerical conditions. On the other hand, the analysis of children's individual responses shows a failure to respond along the extent of the response bar. Many of our preschool children did not spontaneously map numbers onto space, but instead tended to restrict their responses to the bars' endpoints and midpoint locations (bi- and tri-categorical patterns, respectively).

Because all participants experienced a Brightness Testing phase, we have yet to disentangle its possible contribution to the use of the response bar when they are afterwards tested with numerosities. During the testing phase of the brightness pre-training, responses to grays were not differentially rewarded towards a specific mapping; they were only extinguished whenever the anchor positions were chosen. All children responded to locations besides the anchors during the brightness sessions. Although they tended to use the response bar, comparing to adults they preferred to select the two anchors plus the middle portions. It is possible that, when they were presented with the stimuli that differed from the trained ones, an initial response along the bar may have reinforced responding according to the rule "if not anchor stimulus, respond at this location". Perhaps this location may have become a third hallmark and, subsequently, influenced the responses during the numerosities conditions. Within the mental number line hypothesis, it is not clear how performance in the brightness task would impair performance in the numerical task, or even how transfer between psychological scaling would have occurred, especially when considering its statement about an automatic and inherent number-space association. An additional point regarding tri-categorical patterns in numerical tasks is that young children's tendency to concentrate responses at the line's middle portion has also been reported in solely verbal conditions (Barth & Paladino, 2011). Finally, in previous symbolic number-to-line tasks it has not been tested how specific instructions and/or the prior training of the middle numerosity – midline location mapping affect performance.

Another consideration in the current study is the contrast between our use of a computerized task, opposed to prior studies' paper-pencil version and their reliance on response prompting and/or repetition of verbal instructions before each trial. This must have been an intervening variable, though we cannot ascertain its weight. On one hand, prompting and nonvocal cues, such as depicting the anchor stimuli at the bar endpoints, as other studies have implemented, would have helped children to respond (e.g., Demchak, 1990; MacDuff, Krantz, & McClannahan, 2001). On the other hand, it is doubtful that the procedural differences impacted children's motivation, for children did not disengage, a minimum interval of one day between subsequent sessions was kept, and no verbal or nonverbal behavior signaled to the experimenter that the situation was unpleasant for them.

Yet, despite all procedural characteristics, at the usual group-level analysis our results did not differ from previous number-to-line findings and extended them to children's nonverbal numerical representation. However, as we have shown, individual-level analyses made clear that preschooler's performance was highly variable and tended towards categorical patterns. Many preschoolers seem to be judging numerosities' similarity on two or three *manipulanda*, rather than on a continuous spatial response bar. Children's logarithmic group curves are thus averaging artifacts that do not represent individuals' performance. In other words, it is not possible to "induce" from the mean curve to the individual curve (Estes, 1956; Speelman, & McGann, 2013; Trafimow, 2014). Due to this, our findings do not inform us about the psychological scale of children's representation of numerosity. Rather, they show how misleading it may be to take young children's mean response curves (more so group curves) as mirroring a psychological number scale. That average logarithmic curves do not represent individuals' performance has been previously addressed in Núñez and colleagues' study with the Yupno tribe (2012).

Our study adds to the evidence of failure at responding continuously, and of position bias towards the anchor locations. We believe these questions should be extended to previous number-to-line experiments presenting verbal numbers (Arabic digits, spoken words).

In sum, Study 1 revealed that many preschoolers either did not show sensitivity to the numerical sample – as shown by average location not increasing as a function of the numerical sample – and/or did not respond using the entire range of the response bar, restricting their responses to three or two locations. Critically, this categorical mapping was "swept" when averaging the response locations, both at an individual and group level analysis. At the group level, the curves depicting mean response location as a function of numerosity resembled a logarithmic-like pattern, a finding that seemed to align with the hypothesis of a logarithmic-to-linear representational shift (Siegler & Opfer, 2003; Dehaene et al., 2008). However, inspection of single trials revealed that the average should not be taken as a representative of individual responses, let alone providing direct readouts of the psychological representation of numbers (e.g., a mental number line).

If categorical responding is the "a priori" response pattern, a more interesting question would be how a logarithmic pattern is linearized with schooling, and under which conditions (prior learning experiences) does responding in a continuous one-dimensional space first becomes under the control of the (continuous) numerical stimulus dimension.

CHAPTER III : STUDY 2

Control of a response continuum by the numerical stimulus continuum: the effects of pre-training on a non-numerical continuum

1. Introduction

The current study further explores the performance of preschoolers and adults in number-to-line tasks. Preschool children and adults were tested in non-symbolic number-to-position tasks, with the numerical samples presented in the form of simultaneous arrays of 1-to-9 or 10-to-90 dots (conditions Dots[1,9] and Dots[10,90], respectively). Our overall objective was to characterize individual performance. Additionally, we were also interested in inspecting the data both at an average group, individual average and, lastly, individual response levels. This gradual narrowing was meant for discussing possible discrepancies between the conclusions yielded from each unit of analysis.

Also regarding data analysis, in previous studies the authors inspected either the mean or the median response location. For instance, in the works of Siegler and colleagues, it was customary to fit different models to the group and/or individual median curve and compare the fittings based on their goodness of fit (R^2 coefficients). Experimental groups and/or participants were then classified as cases for the model that yielded the highest R^2 value (Siegler & Opfer, 2003; Siegler & Booth, 2004; Barth & Paladino, 2011). Conversely, in their number-to-line experiments with the Mundurucu participants (with numerical samples in the formats of Dots[1,10], Tones[1,10], and Numerals[1,10]), Dehaene and colleagues (2008) opted to inspect mean response location. However, when studying western children with Arabic digits 1-100 and 1-1000 number-to-line tasks, they chose to describe the median response location (Berteletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010). In their set of experiments investigating how adult Americans mapped sets of Dots ([1,10] and [1,00]), Tones ([1,10]), and spoken number words [1,19] onto spatial (positions in a number-line) and nonspatial reporting conditions (intensity of squeezing a dynanometer, striking a cowbell and vocalizing), Núñez and colleagues (2011) reported group mean responses. Later, when studying how participants from an indigenous tribe – the Yupno from Papua New Guinea – mapped symbolic and non-symbolic numerosities into space (Dots[1,10], Tones[1,10], Numerals[1,10]), Núñez and colleagues chose to inspect median response location (Núñez, Cooperrider, & Wassman, 2012).

To our knowledge, thus far no study has commented on the consistency between the two average measures. We will do so in the current study, and try to relate such consistency with categorical patterns of responding. It is not clear, however, if one of the central tendency measures should better reveal categorical responding. On one hand, at the individual curve level, when a participant is sensitive to number but responds at the two anchor sites in different proportions, the curve connecting the medians, less influenced by outliers, should be steeper than the mean curve. On the other hand, we are not sure whether such effect would be seen at the group level, especially if only a small proportion of the participants respond categorically, while the others respond continuously. To exemplify, if participant "A" responds preferably at the smaller anchor location following the first sample, and preferably selects the larger anchor location following the remaining samples, average response location will rise abruptly and stabilize at a asymptotic maximum (i.e., plateau) from the second numerical sample and onwards. But if participant "B" preferably responds at the smaller anchor location following the first three smaller numerosities, and at the right endpoint following the remaining numerosities, then participant "B"'s average curve would plateau at greater numerosities than participant "A"'s curve. Whether or not the group averaging between these and others' individual curves would reveal categorical-like features would depend on the balance between the number of "categorical" participants and how many times each numerosity is sampled (i.e., number of trials for each numerosity).

Moreover, we have pointed out that the unit of analysis in former studies has been the group curves, which in most cases involves the comparison between different age, schooling or disability groups. Perhaps because the authors did not intend to address individual differences, each participant provided only <u>few responses for each numerical sample</u>. For example, in the study with the Mundurucu tribe, Dehaene and colleagues (2008) asked the participants to provide two responses per numerical sample. Similarly, in their number-line experiments with Arabic digits, Opfer, Booth and Siegler typically tested their participants in one or two trials per numeral (1 in Siegler & Opfer, 2003; Opfer & Siegler, 2007; Booth & Siegler, 2006, 2008; 2 in Siegler & Booth, 2004; Siegler & Ramani, 2009). And possibly because they were interested in direct data comparison, in studies that followed, the procedure entailed just few responses per numerosity as well (e.g., Muldoon et al., 2011; Ashcraft & Moore, 2012). In our previous study, each participant provided four responses for each numerical sample. But, since in the current experiment data collection was projected to discuss the results at the individual unit of analysis, each participant provided 16 responses per numerical sample.

Rationale for the methods of the current experiment

Consider the number-to-position task from the standpoint of stimulus control studies. One way to describe the procedure is as a continuous-response task, in the sense that "the set of available response alternatives and the set of outcome values are each ordered along a spatial or other clearly specified continuum" (Rosenberg, 1963). In other words, it asks for a continuous response dimension – responding along a line - , to be under the control of a continuous stimulus dimension – number magnitude. All number-to-position procedures entail an initial training of the two anchor points: smallest number - leftmost position and largest number –rightmost position. A testing phase follows, during which novel, intermediate numerosities are also presented.

Interestingly, in human and animal experiments with other stimulus dimensions, training with only two points would hardly lead to untrained stimulus points evoking untrained, intermediate responses (Wildemann & Holland, 1972; Scheuerman, Wildemann, & Holland, 1978; Stoddard & McIlvane, 1989). Likewise, based solely on anchor training, one might expect the nonsymbolic number-to-position task to prove hard to accomplish for a young child, let alone to spontaneously occur when they are first presented with the task. Hence, the idea of an inherent spatial number line or, in other words, that humans would spontaneously (i.e., without explicit training) respond continuously, seems to be at odds with how difficult it appears for a continuous repertoire to be established in comparison with, say, a dichotomous response such as selecting between two response *manipulanda* (or, in psychophysics terminology, a two-alternative forced choice (2AFC) procedure; Fechner, 1889).

Nevertheless, even severely impaired children may present continuous repertoires. And even if spatial mapping is not a predisposed or a "hard-wired" ability, young children are likely to have already experienced other continuous repertoire situations. This was the reasoning for the introduction of a Brightness-to-position task prior to the testing with number-to-position task. We expected that the formation of a continuous repertoire between increasingly darker stimuli and ordered positions along a response bar would result in "savings" or transfer of learning (Mazur, 2002, p. 242) when children are asked to order numerosities along a response bar.

Unfortunately, because prior to being tested in Number-to-position tasks all participants were tested in a Brightness-to-position task, in the previous experiment (Study 1) it was not possible to ascertain the specific effects of this task. For that reason, in the current experiment we tested the effect of a pre-training with a Brightness-to-position task by separating participants into Control and Experimental groups. The Control Group was tested solely in Number-to-Position tasks, but the Experimental group received a pre-training with the Brightness-to-position task. However, some changes were introduced with respect to Study 1. In the previous experiment, we did not ensure that participants had learned to map brightness stimuli onto space; we merely instructed them to do so and then tested them. In the current experiment we ensured that the participants in the Experimental Group were trained in a continuous response topography (Wildemann & Holland, 1972).

There were a few other alterations implemented in the Brightness pre-training procedure. Anecdotal reporting from the previous procedure suggests that some preschoolers occasionally responded disregarding numerosity because the sample was not present at the response moment. This is probably an instance of the somewhat trivial finding that even when a subject must select a choice that equals the sample (Identity Matching To Sample - MTS), introducing a delay between the sample presentation and the choice moments (Delayed Matching To Sample - DMTS) decreases accuracy (Roberts, 1998, p. 72; Mazur, 2002, p. 257; Chelonis et al., 2000, 2014). Moreover, in Study 1 the Brightness and the Number dimension tasks did not proceed as similarly as possible. Our former Brightness-to-position pre-testing resembled a DMTS task in the sense that the sample stimulus (a colored square) remained on the screen for few seconds, then it disappeared and the comparison stimuli (a bar forming a greyscale) were presented. When it was time for the subject to respond (i.e., touch at a location within the colored response bar), the sample was no longer present. Note, however, that the correct location on the response bar had the same color as the sample stimulus, so sample and choice corresponded directly. Our former Number-to-Position task, however, was similar to a conditional discrimination or symbolic (arbitrary) DMTS, for the sample and the choice stimuli were completely different (Mazur, 2002, p. 257). Namely, the response bar always had a uniform white background, and the subject had to respond based on location only.

In the current experiment, a better parallelism between the brightness- and number-to-position procedures will be implemented by having a response bar uniformly filled in yellow during both tasks, and by having the sample stimulus present during the choice moment, so that both tasks resemble symbolic MTS procedures.

Another procedural difference will be the substitution of the mouse device by a touchscreen. Touchscreens are known to be easier for novices to learn and have been shown to be a more effective input device for preschool children (Lu & Frye, 1992; Hourcade et al., 2004). The main reason is that it allows direct pointing, that is, a direct overlap of the touchscreen space and the display screen. Moreover, while it was not the case in our Study 1, some computerized number-to-line studies have relied on the experimenter to operate the mouse device. In those studies, for example, children would point with their finger to the line site and the experimenter moved and clicked the mouse pointer at the selected location (e.g., Landerl, Fussenegger, Moll, & Willburger, 2009; Dehaene et al., 2008). Thus, with the touchscreen input device and the simultaneous presentation of sample and choice stimuli, we expect our procedure to become easier for child participants and reduce unintended responding.

Finally, we present the last aim of the current study, which is to better understand how number-to-position performance relates to other numerical abilities. A pervasive idea in the number mapping literature, present ever since the first number-to-space study, is that children represent the smaller, familiar numbers linearly and the larger, unfamiliar ones, logarithmically. In addition, the linear- and, conversely, logarithmic-like growth of the estimates (responses along the line) is conceptualized one step further, that is, as a property of the representational system or "mental number line". Lastly, how linearly numbers are mapped onto space is supposed to depend on the child's familiarity and exposure to numbers in different ranges (Ebersbach et al., 2008), as well as on formal math knowledge (Booth & Siegler, 2006; Goksun et al., 2013).

The first study that tested kindergarten children in number-to-position tasks (Siegler & Booth, 2004, Exp. 1) also investigated the relationship between mapping performance and formal math knowledge. Linearity in number-to-position tasks was measured by how well the data was described by a simple linear model, as measured by R^2_{lin}. A statistically significant but moderate correlation was found between R^2_{lin} scores and the math achievement test scores (mathematics section of the SAT-9 -

Stanford Achievement Test). In other words, the more linear a child's number-to-position estimates, the higher his/her math achievement test scores. Since this study, other authors have found statistically significant, low to moderate, correlations between linear mappings and formal math knowledge (Booth & Siegler, 2006, 2008; Berteletti et al., 2010; Muldoon et al., 2011; Ashcraft & Moore, 2012; Sasanguie et al., 2012).

However, amongst the two variables proposed as foundations for a linear psychological representation, studies relating number-to-position performance with scores from math achievement tests quite outnumber those who address familiarity with the numerical range. To our knowledge, the first study to empirically address the effect of familiarity with numbers was that of Ebersbach and colleagues (Ebersbach, Luwel, Frick, Onghena, & Verschaffel, 2008). In this study, children from kindergarten to 3rd grade (n ≈ 20 for each grade) were tested in a 1-100 or a 1-1000 number-to-position task, where the numerical sample was a card with a printed Arabic numeral and a corresponding number of dots. Each trial started with the experimenter reading this number and presenting the child with a hypothetical story on how many chocolates would be needed for *X* guests attending a birthday party, where *X* was the numerical sample. The child responded by moving a pointer along a wooden stick to indicate the spatial estimate. The session ended with a counting evaluation, which in turn was adapted from a previous experiment by Lipton and Spelke (2005). In this counting assessment, children had to complete a series of decade transitions, such as completing the number word series following "55, 56, 57" until the child completed the decade change. Ebersbach and colleagues found a moderate correlation (r = .31) between the number-to-position performance and the abstract counting performance. The fact that they did not carry separate correlations for each age group, however, gives rise to cautionary thinking when extrapolating the role of familiarity with the numerical range.

In Berteletti and colleagues (2010)' study, kindergarten children were tested in symbolic 1-to-10 and 0-to-100 number-line tasks (Experiment 1). A small but significant correlation was found between the performance on the smaller and larger range conditions. The authors interpreted this correlation as an evidence for the log-to-linear shift in numerical mental representations. To recall, this hypothesis defends the notion that the transition between numerical representations occurs by experiencing information that does not match the earlier logarithmic representation, and this learning progresses first in a familiar smaller numerical context, before being applied to larger numerosities (Opfer & Siegler, 2007; Berteletti et al., 2010). In this sense, the more accurate (i.e., linear) a mental representation in the smaller range, the more likely is the child to extend this representation to larger numerosities. However, in their first experiment, Berteletti and colleagues (2010) did not test counting and familiarity with numbers, a feature that they corrected in their second experiment. In their second experiment, 373 kindergarters, from 3.6 to 6.3 years old, completed 1-to-10 and 1-to-20 number-to-position tasks – which are considerably closer numerical ranges than those in their first experiment. Afterwards, each child was scored on his/her ability to count their fingers up to 10, or to order Arabic digits or sets of dots in the [1,5] range. In this experiment, instead of addressing linearity directly by the R^2_{lin} measure, the authors correlated each child's type of representation on the number-line task (None = 0, Logarithmic = 1, and Linear = 2) with the counting abilities. A significant and weak to moderate correlation was found between representation and the ordering tasks, but no correlation was found between the type of representation and the finger counting task. Therefore, the evidence remains ambiguous as to the role of counting and familiarity with numbers being "a prerequisite to apply a linear strategy" (Berteletti et al., 2010).

In fact, "familiarity with numbers" is a broad term, which certainly includes more than the ability to recite the number words up to a certain limit (Lipton & Spelke, 2005; Le Corre & Carey, 2007; Ebersbach et al., 2008). Adding to that, in children close to our subjects' age, discrepancies can be found between the different components of numerical abilities (e.g., Dowker, 2008). Given the scarce empirical evidence to back the theoretical view that familiarity with numbers is responsible for the log-to-linear transition, the last phase of our study would assess children in a set of estimation and counting tasks. In addition to abstract counting, we intended to evaluate serial counting of objects, relations between objects and number words (e.g., "Give-N" tasks) as well as a nonsymbolic estimation (sets of dots). We planned to find which, if any, of these counting and estimation abilities correlate with performance in the nonsymbolic number-to-position performance.

2. Method

2.1 Participants

The experiment studied twenty-four Portuguese pre-schoolers (12 girls) and twenty-four Portuguese young adults (12 women). The mean age of preschool children was 5.29 years (SD = 0.59; range 4.06 - 6.23) and the mean age of adults was 21.27 years (SD = 2.51; range 18.13 - 25.69). Informed consent was given by all adult participants and the children's parents.

2.2 Numerosity stimuli

Two databases of numerical stimuli were constructed, one for the smaller range condition (Dots[1,9]), the other for the larger range condition (Dots[10,90]). Within each database, the numerical stimuli consisted of unique sets of red dots (RGB: 242, 0, 60) presented against a green background (RGB: 183, 218, 177). Individual dot position was randomly determined, within an invisible matrix of 144 (12 * 12) possible positions. This invisible square grid was inscribed on a circle (diameter 7 cm), with the same green background.

All dots in a set had the same size, but individual dot areas of the sets presented during the Training phase differed from those in the Testing phase, as per Dehaene and colleagues (2005)' suggestion regarding the implementation of controls for non-numerical parameters in the creation of numerical stimuli (Figure 1). Namely, dot area varied during the Training phase, so that the sets for the "Many" anchor and for the "Few" anchor were matched on total occupied area. The parameters for the sets to be presented during the Testing phase - sets comprising the "Few" and "Many" anchors and seven intermediate numerosities – depended on the previous ones. During Testing individual dot area was fixed and now the total surface area increased with numerosity.

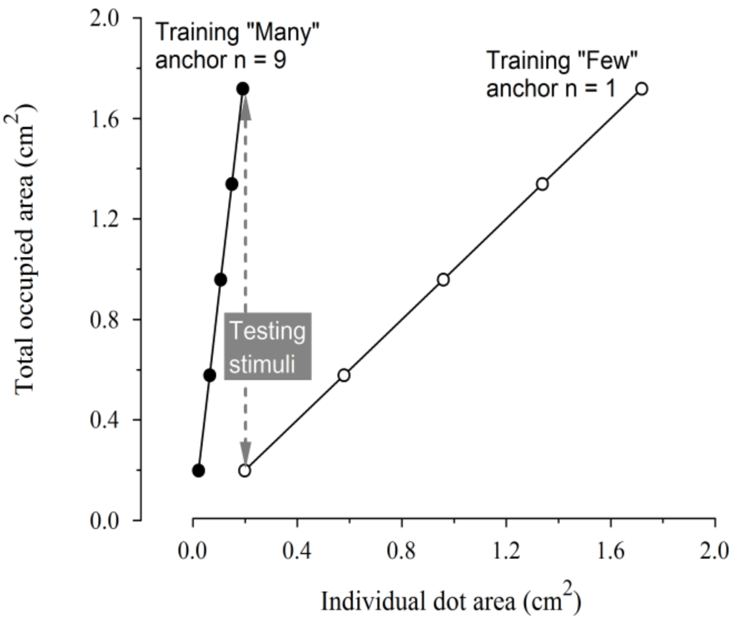


Figure 1. Implementation of controls for non-numerical parameters when creating the numerical stimuli from the smaller range (Dots[1,9]) database. During Training, individual dot area was varied so total occupied area in the sets with the “few”-anchor number of dots were matched with the total area occupied by the “many”-anchor. During Testing all numerical sets had the same individual dot area, but total occupied area increased as a function of numerosity.

An additional control was that the distribution of the parameter that was statistically fixed during Training – total occupied area – was chosen so that the smallest value (0.20 cm^2) is equal to the smallest value presented in the Test phase, and the largest value (1.8 cm^2) in Training is equal to the largest value presented in the Test phase (Figure 1). This way, all the summed area values presented in the Testing phase would be presented in the Training phase, so that all Testing sets are equally non-novel with respect to total occupied area.

In Figure 2 we present some examples of the numerical stimuli. For the smaller range database (numerosities from 1 to 9), training “Few” and “Many” sets had an average surface area of 0.96 cm^2 (ranging from 0.20 to 1.8). This required individual dot size to be smaller in the “Many” sets, than in the “Few” sets. Namely, individual dot area ranged from 0.2 to 1.8 cm^2 in the “Few” anchor training sets, and from 0.02 to 0.20 cm^2 in the “Many” anchor training sets. We constructed 32 distinct numerical sets for each of the trained anchor numerosities (total of 64 stimuli).

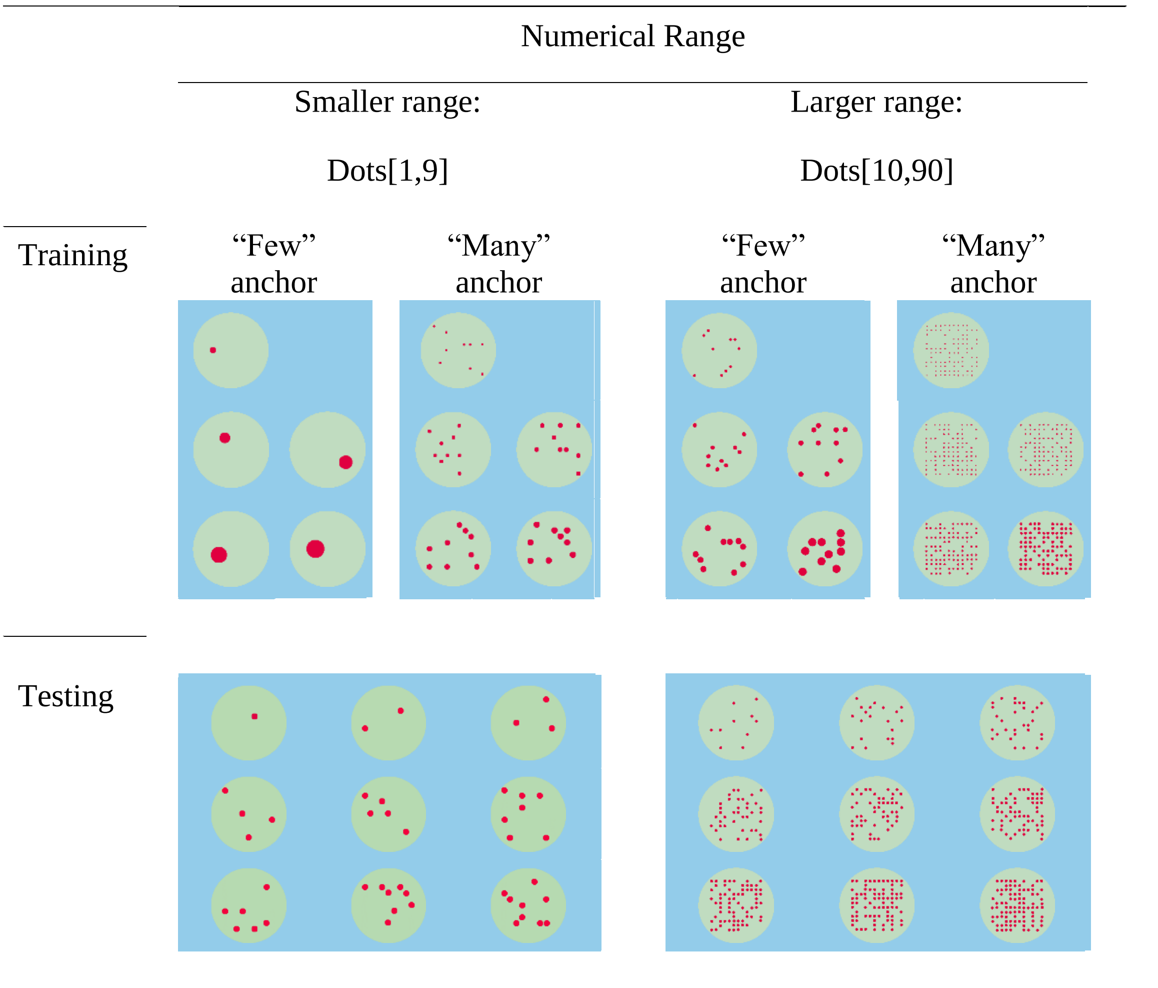


Figure 2. Examples of the numerical stimuli used in the Training and Testing phases, for the smaller (Dots[1,9]) and for the larger (Dots[10,90]) range conditions.

As for the Testing numerical stimuli, we created 16 exemplars for each of the integers between 1 and 9. All had element area equal to 0.20 cm^2. For the larger range condition database (dots from 10 to 90), the relation between individual dot size and total occupied area remained the same in the Training and Testing phases. Therefore, element size of the training stimuli for the "Few" anchor (10) ranged from 0.2 to 1.8 cm^2, which means that average surface area ranged from 2.0 to 18 cm^2. Because training anchors were matched on average surface area, element size for the "Many" anchor (90) ranged from 0.02 to 0.2 cm^2.

2.3 Procedure

Participants were seated in front of a touchscreen laptop (HP Pavilion tx2000 Notebook PC, screen size 12.1", screen resolution 1024 x 768, refresh rate 60 Hz), in a quiet room of the school. A program written in Visual Basic controlled all session events and recorded the participants' responses. The experimenter remained in the room, seated about 0.75 m behind the participant to keep out of his sight and prevent response bias. A separate monitor, positioned behind the participant and facing the experimenter, was connected to the laptop and displayed the experimental events.

Each boy was paired with a girl of the closest age. We then selected the two most proximate (in terms of age) boy-girl pairs and these four participants were distributed into the two experimental groups by selecting one participant and tossing a coin. The placement of the first participant meant that the other member of the pair (of the opposite sex) went to the same experimental group, whereas the members of the other pair went to the other experimental group.

The experimental conditions and, consequently, the number of experimental sessions, depended on the experimental group of the participant. Participants in the Control Group were tested solely in a numerical line mapping task, whereas those in the Experimental Group were trained in a line mapping task with brightness stimuli prior to being tested in the numerical mapping task. In the end of the experiment, all participants were assessed in a set of counting and verbal estimation tasks (Figure 3).

1. Brightness to Position (Pre-training). During the first experimental phase, participants in the Experimental Group learned to respond in a spatial *continuum* as a function of a non-numerical stimulus dimension, brightness.

1.1. Training two anchors. At the beginning of the session, participants were presented with the response bar, which was a uniformly colored yellow rectangle, 26 cm wide x 1.4 cm high, located 12.5 cm below the upmost part of the screen. The experimenter instructed that the response bar was a path that went from the least dark to the darkest color and, while saying that, moved her finger along the figure. Next, she asked the child to also move his/her finger from the beginning up to the end of the path. The experimenter added that the child would see circles and that as the circles got darker he/she would have to touch the bar to increasingly rightmost positions.

Participants were told that they would start by learning what stimuli belonged to the start and end positions of the path, then began the actual experiment.

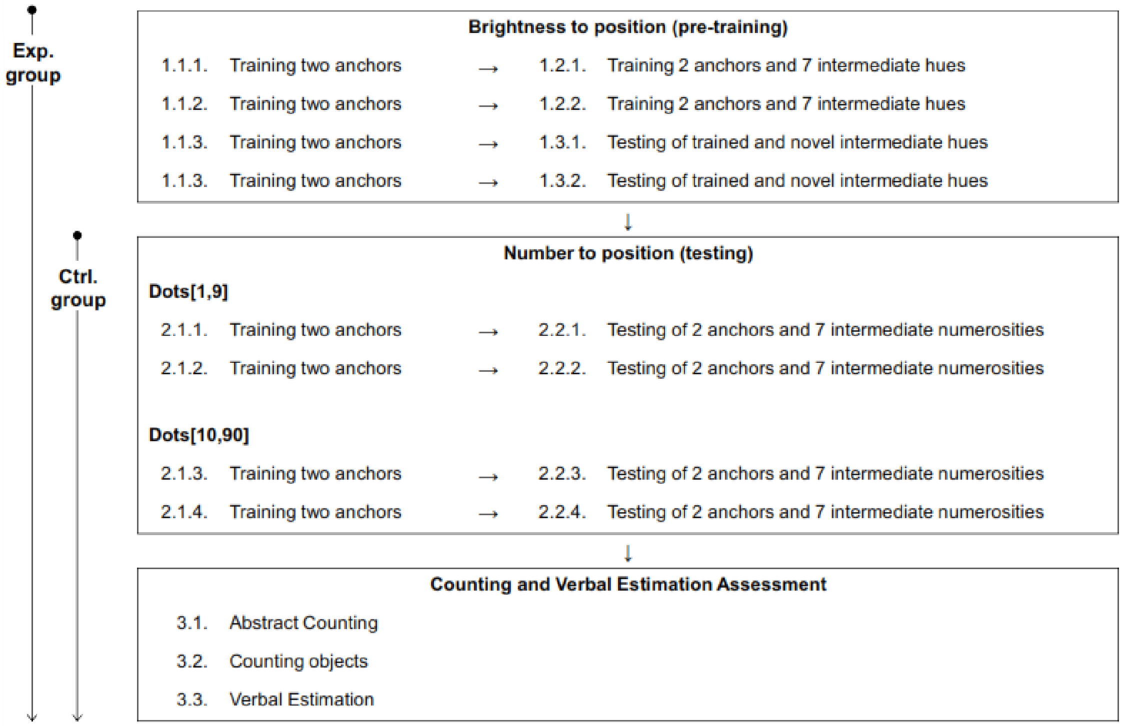


Figure 3. Diagram depicting the progress through the three experimental phases: Brightness to position (pre-training), Number to position (testing) and Counting and Verbal assessment. Participants in the Experimental group underwent the Brightness to Position pre-training before being tested in the Number to Position tasks, whereas Control Group participants started from the Number to position tasks onwards.

Each trial started with the screen completely colored in cornflower blue (RGB color: 147, 204, 234), on which, following an Inter-Trial-Interval (ITI) of 1.5 seconds, a yellow star image (diameter about 3 cm) appeared at a random location. A touch to the star triggered the appearance of the yellow response bar and of a sample stimulus. Sample stimuli were uniformly colored circles (diameter 7 cm), either white ("least dark") or black ("most dark"), displayed horizontally centered and located about 0.1 cm below the upmost part of the screen. Both the sample stimulus and the response bar remained on the screen until either a touch occurred at any location of the bar, or 20-s had elapsed, whichever happened first.

A touch response along the bar was visually signaled by a thin dark blue vertical strip (3 pixel thick) located at the x coordinate of the response, splitting the response bar into two parts. No other touchscreen responses were visually signaled and the mouse cursor remained invisible throughout the experiment. The program saved choice latencies and the x/y coordinates within the response bar. If the sample was white (RGB color (255, 255, 255)), the correct response was a touch at the leftmost portion of the bar, whereas the rightmost portion of the bar was correct after a black sample stimulus (RGB color (0, 0, 0)). The acceptable region was within 1/9 of the bar length.

Following a correct response the experimenter personally gave the verbal feedback "*You did well, that is the correct place*", then touched the selected location, thus terminating the current trial and starting a new one. If the response was incorrect, the experimenter gave the feedback "*That was not the correct place. Let us try again*". A correction procedure started, which consisted of repeating the trial twice. In the first repetition, all trial events were similar to a regular trial except that at the response moment the experimenter held the child's hand and moved it along the bar, from its leftmost portion up until the correct location, saying "*When you see that, you must follow the path up until here*". Where to respond was previously signaled to the experimenter in the monitor placed behind the participant. In the second repetition, the participant responded on its own, as in a regular trial; if this response was correct he received the positive verbal feedback and advanced towards the next trial; if incorrect, the correction procedure was re-started, maintaining the two-step loop of a first experimenter-guided trial followed by a self-directed response trial. The "putting through" correction technique was selected because it proved to be successful in establishing an experimenter's response as a discriminative stimulus for the same response by the subject, even in profoundly intellectually disabled children (Konorski & Miller, 1937; Baer, Peterson, & Sherman, 1967).

Training progressed until the participant: (i) completed at least twenty or ten trials (depending on it being the first training with the anchors or not), not counting the repeated trials and (ii) after he had reached the criterion of five consecutive correct responses for each anchor stimulus. Adult participants, though, always ended the training after reaching the criterion of five consecutive correct responses for each anchor stimulus.

1.2. Training two anchors plus seven intermediate hues. Following the acquisition of the two anchor mappings, and still in the same experimental session, the participants were told that, in addition to the least dark and the darkest stimuli, they would now also see other values that ranged from one end of the bar to the other end. The experimenter informed that the task was to decide where to place the sample along the response bar, attending that stimuli got "darker" as one moved from the left to the right of the response bar. While providing this instruction, the experimenter was touching the bar, starting from its left endpoint and progressively moving towards the right endpoint.

The trial events were similar to training trials except that the sample stimulus (i.e., the circle's color) could have one of nine brightness values, the previous two anchors and seven intermediate grays. More specifically, sample stimuli were a consistent grayscale group of nine patches from 255 (white) to 0 (black) in RGB (in other words, a grayscale from 0% to 100%). Correspondingly to the nine stimulus hues, there were nine acceptable response areas of the bar. Response areas for the white and black stimuli were as in the previous anchor training phase. For the second, up until the eighth hue samples, the width of the acceptable region included 1/9 of the response bar length plus an admissible overlap of 20% between adjacent portions. These areas were not observable to the participant since the response bar was always uniformly colored in yellow.

The correction procedure was in effect whenever an incorrect response occurred. Participants underwent two sessions of training with the nine hues, such that each session was always preceded by the training of the two anchors (1.1). Each session comprised eight exemplars of each hue, resulting in a total of 72 trials per session. The order of stimulus presentations was randomized across trials. There was at least a day of interval between consecutive experimental sessions. At the end of each session, the child participant received a sheet with stickers.

1.3. Testing with previously trained and novel intermediate hues. After learning to map the nine hues along the response bar, participants in the Experimental Group were tested with seventeen hues, namely the nine previously trained plus eight new interpolated grays. Differently from the previous sessions, no feedback followed the response, and no correction procedure was in effect. A response was signaled by the blue vertical line bisecting the bar at the selected location for 750 ms, after which the 1.5 s ITI started.

Participants underwent two sessions of testing with the seventeen hues, and each session comprised 72 trials: 6 of each anchor (white and black), and 4 for the remaining greys. There was at least a day of interval between consecutive experimental sessions and at the end of each session the child received a sheet of stickers.

2. Number to Position. Participants in the Experimental Group began the Numerosity Testing Phase after an interval of at least a day since their last testing session with brightness stimuli. For participants in the Control Group, Numerosity Testing was their first session. Participants were tested in two ranges between the few and the many value, namely the smaller range, Dots[1,9] condition (numerosities from 1 to 9), and the larger range, Dots[10,90] condition (numerosities from 10 to 90). Order of numerical range was counterbalanced across subjects.

2.1. Training the numerical anchors. Each numerical session started with the training of the correct location of the anchor stimuli. The experimenter told the participant he/she would start a game with numbers. The response bar was (now) a path that went from the less numerous to the most numerous locations and, while saying that, moved her finger along the bar. The experimenter invited the child to point at the position where the lesser number of dots should be. Then, he told the child to move his/her finger progressively until the end of the path and as he/she did so, explained that the more dots there were the further along the path the child should touch. Next, the experimenter said that they would begin by learning about the numbers belonging to the start and end positions of the path (refer to Appendix A for the complete transcription of the verbal instructions preceding the Number-to-Position computerized procedure).

All session events were similar to those described for the Brightness Training Phase, including the ITI, the subject's response to start the trial, the spatial and temporal configurations of sample and choice stimuli, the yellow response bar, and the feedback contingencies. The difference was that, instead of a uniformly colored circle the sample was a numerical stimulus. During this phase, numerical stimuli were the Training stimuli previously described (Figure 2), either sets of 1 and 9 dots (Dots[1,9] condition), or 10 and 90 dots (Dots[10,90] condition). Anchor training continued for at least 20 trials and until five consecutive correct responses occurred for each anchor.

2.2. Testing anchors and intermediate numerosities. Following anchor training, participants started the Testing Phase. They were told they would see several numbers of dots, ranging from the minimum to the maximum number of dots. The experimenter told the child that, knowing that the minimum number of dots was at the beginning of the path and the maximum number belonged at the path's end, they would have to decide where to place each number of dots along the path. She moved her finger progressively until the end of the path and reminded the child that the more dots there were the further along the path the child should touch.

The testing phase comprised 72 trials, with eight unique set configurations for each numerical sample (the two anchors and the seven arithmetically spaced intermediate numerosities), presented at random. During testing, no verbal feedback followed a response. After the subject touched the response bar, the dark blue line appeared bisecting the bar and stayed on screen for 750 ms, and then the ITI started.

There were two numerical sessions for each numerical range, interspaced with a day between them. All numerical sessions began with the training of the two anchors before advancing to the testing with all the numerical values. Thus, within each numerical range, a subject emitted 144 test responses (8 trials * 9 numerosities * 2 sessions).

3. Counting and Verbal Estimation Assessment. To examine how performance in number line tasks correlates with proficiency in counting and verbal estimation, in the last experimental day participants were screened in a series of counting and verbal estimation tasks in the case of children, or only the latter task in case of adults.

3.1. Abstract Counting. This procedure was adapted from one of the counting tasks in Lipton and Spelke (2005)'s study with pre-schoolers. The experimenter asked the child which was the higher number he/she could count up to and wrote down the answer. Alternatively, if the child spontaneously started counting, the experimenter wrote down the stream of numbers.

Based on this answer, the experimenter started to count and the child had to continue this sequence after the experimenter stopped. The sequences were selected to test for decade transitions (e.g., "continue after me: 56, 57, 58, …"). If the child failed the transition, the next sequence would include the preceding decade change (e.g., "46, 47, 48, …"). If the child managed to respond correctly, the next sequence involved the subsequent decade change (e.g., "55, 56, 57, …"). If the child failed even the smallest decade transition, the experimenter asked her to start counting from 1 up, and registered the biggest number until an error occurred.

3.2 Counting objects. Children were also assessed on their serial counting abilities (see the experimenter's recording sheets in Appendix G).

3.2.1. Serial marble counting. Children were presented 9 marbles aligned horizontally and were asked to count aloud the marbles and touch each of them as they were counting. The child did not see the experimenter aligning the marbles before presenting them. Besides counting proficiency, we were interested in seeing whether children spontaneously showed evidence of a directional, leftward or rightward, bias.

3.2.2. Correspondence between marbles and numerals. The following tasks were adapted from those in Opfer and Thompson (2006)'s, Wynn (1990)'s, and Huttenlocher, Jordan and Levine (1994)'s procedures. Children were asked to match between number words and marbles or between marbles and marbles. In each task they were tested in all numerosities, from 1 to 9. The order of presentation of numerosity, as well as of task, was counterbalanced across participants using a Latin Square design.

3.2.2.1. Numerals to Marbles (also named "Give-N" tasks). The experimenter gave the child a pile of 25 marbles. Next, he asked the child to give him n marbles, with n = 1, 2, …, 9.

3.2.2.2. Marbles to Numerals. The experimenter presented {n} marbles and asked the child to say how many marbles there were.

3.2.2.3. Marbles to Marbles. The experimenter presented {n} marbles and asked the child to take from his pile of marbles the same number of marbles he was being presented with.

In the last two tasks, the experimenter presented sets of marbles to the child. These sets of marbles were assembled out of the child's view, to avoid the possible cueing effects of the experimenter's motions of putting each marble one after the other (e.g., serial counting).

3.3. Verbal Estimation of sets of dots. The last task for children participants, and the only assessment task in the case of adult participants, required for them to verbally estimate the number of dots in a simultaneous set.

Our verbal estimation procedure was adapted from Lipton and Spelke (2005)'s study with pre-schoolers. One difference was that, while in their study each set of dots was presented too briefly to be counted, in ours the set remained on screen until a verbal response or after 10 seconds had elapsed, whichever occurred first.

First the experimenter presented two sets, one depicting 1 dot and other depicting 9 dots, and enumerated them accordingly. Next, participants were told they would see many images containing between 1 (pointing to the image with one dot) and 9 dots (pointing to the image with nine dots) and their task would be to say how many dots they thought they were seeing. The experimenter instructed the participant to try and avoid counting the dots, and give his/her "first impression" on the number of dots presented. If and whenever the participant showed to be counting the dots, the experimenter did not intervene but registered it in the recording sheet (Appendix G).

The session comprised 36 presentations, with 4 exemplars of each of the numerosities from 1 to 9. In case of incorrect estimates, the experimenter only provided verbal feedback if the stimulus had been a set of 1 or 9 dots, or if the participants' estimate was larger than 9. In those cases, an image of each anchor was depicted and enumerated and the participant was reminded that he was seeing sets of dots within that range. Next, the trial was repeated.

After being tested in this smaller range, the same procedure was implemented in a second session, in tens from 10 to 90.

3. Results and discussion

3.1 Brightness Pre-training

All child and adult participants in the Experimental Groups learned the anchor mappings of white-leftmost and black-rightmost as well as the trained locations for the seven intermediate grays. Children required on average 21correction trials during the first session of the **1.2 Training phase**, and 18 trials during the second session of training. Adults required fewer correction trials, on average 9 in the first, and 4 in the second session. Figure 4 depicts the mean response locations as a function of the grayscale exemplars. The cross data points depict the performance during these two training sessions with two anchors and seven intermediate grays. Correction trials were not included in the analysis.

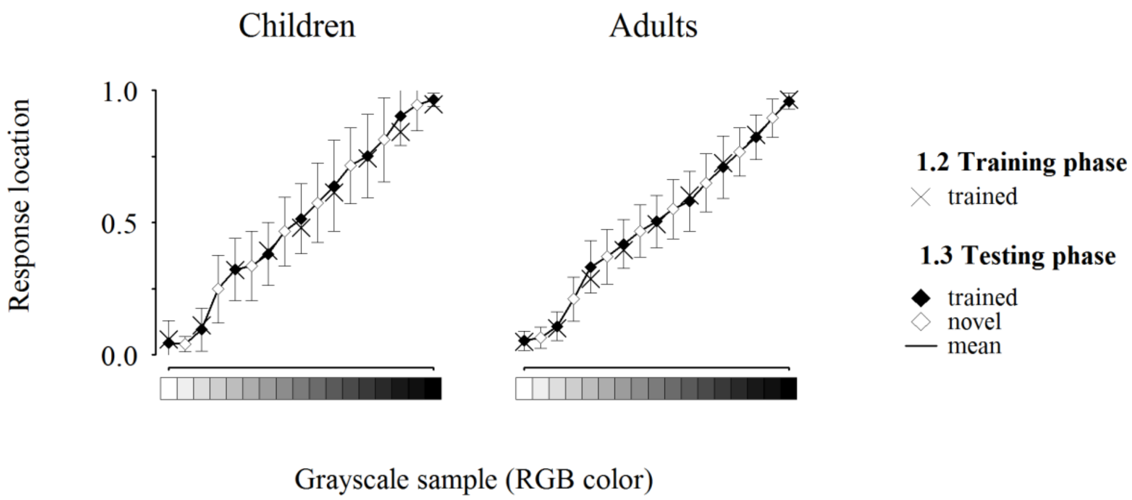


Figure 4. Response locations during the Brightness Training and Testing phases for the children and adults' Experimental Groups. The cross data points depict performance during the 1.2 Training phase, the diamond symbols during the 1.3 Testing phase. Filled diamonds represent the mean response location (± 1 SD) for previously trained sample hues, the white diamonds the mean of novel, interpolated hues.

Participants were successfully trained to respond along the bar when presented increasingly darker stimuli. Moreover, they were successfully trained to keep a constant spacing between consecutive sample stimuli, for mean response location increased linearly with the sample (significant positive Pearson's correlation coefficients of 1 in the Children, $r(7) = 1$, $p < .001$, and Adults, $r(7) = 1$, $p < .001$, groups).

The main interest was to see whether the control of the stimuli hues over the spatial responses established during the training phase was maintained during the generalization test with former and also novel, interpolated hues (i.e., during the **1.3 Testing phase**). Again, a significant positive and linear association occurred between responses along the bar and the sample's brightness (significant Pearson's correlation coefficients of 1 in the Children, $r(15) = 1$, $p < .001$, and Adults $r(15) = 1$, $p < .001$, groups). In Figure 4, the test results are depicted by the diamond data points. Filled diamonds depict responses during testing to stimuli that had been previously trained and unfilled diamonds the novel hues. Recall that during the test phase, neither responses to both the nine previously trained hues and the novel eight stimuli were followed by feedback, nor was the correction procedure in effect. Superposition between the filled diamonds and the cross datapoints displayed in Figure 4, confirms that participants maintained the trained mappings. Additionally, the white diamonds data points are located at locations in-between consecutive trained values. The error bars shown in Figure 4 are smaller for the first two stimuli and the black anchor but remained approximately constant across changes in brightness between the grey stimuli ($SD \sim .13$ in children, $SD \sim .10$ in adults). This indicates an equal dispersion of individual responses, for each grey stimulus, be it a previously trained or a novel stimulus. Finally, visual inspection of each subject's scatterplot depicting single trial responses confirms that all participants in the Experimental Group, children and adults, responded along the bar as a function of increasingly darker stimulus hues (see Appendix B).

A successful performance on this pre-training was important because, firstly, it established that children could complete sessions with at least 72 trials, without becoming tired or stopping to respond. As such, the same number of trials would be presented to them during the numerosity testing phase. Secondly, this pre-training phase in the Experimental group showed that the participants could learn to map two continuous dimensions. Our hypothesis was that having to master this pre-training, particularly for children, would lead to a larger number of participants responding in a continuous pattern when presented numerical samples, in comparison with those participants who had not received this pre-training (Control Group).

3.2 Numerosity Testing

3.2.1 Group analyses

All participants learned the anchor mappings. For both children and adults participants, response locations increased with numerosity. In the left portion of Figure 5, we depict the data from the Experimental and Control groups from adult participants; the right side of the figure depicts children's results. Within each age group, in the upper row we present the data from the smaller range Dots[1,9] condition, and in the lower row from the larger Dots[10,90] range condition. A further separation within each age group is the y-axis in the left graph depicts the group's mean location of responses, and the right graph the median. This inquiry is of special interest due to the weight of the group curve on the discussion and proposal of models. If the two central tendency measures offer different implications for the testing of models, then a clarification has to be made on how they relate.

Adults. The trained anchor mappings were preserved during testing: adult participants responded at the leftmost position when presented the smallest numerosity and at the rightmost position when presented the largest numerosity. As the almost straight lines in the upper row of Figure 5 suggest, adults presented with sets from 1 to 9 dots placed them linearly, with subsequent numerosities placed along the bar with approximately equidistant spacing between them. Due to these two aspects, response location as a function of the numerical sample resembles the diagonal line of the scatterplots. That is, a same straight line with slope 0.125 describes adults' mapping of the Dots[1,9] condition, irrespective of the experimental group and central tendency measure (simple linear regression analyses with fixed slope = 0.125 yielded R-squared values of 1).

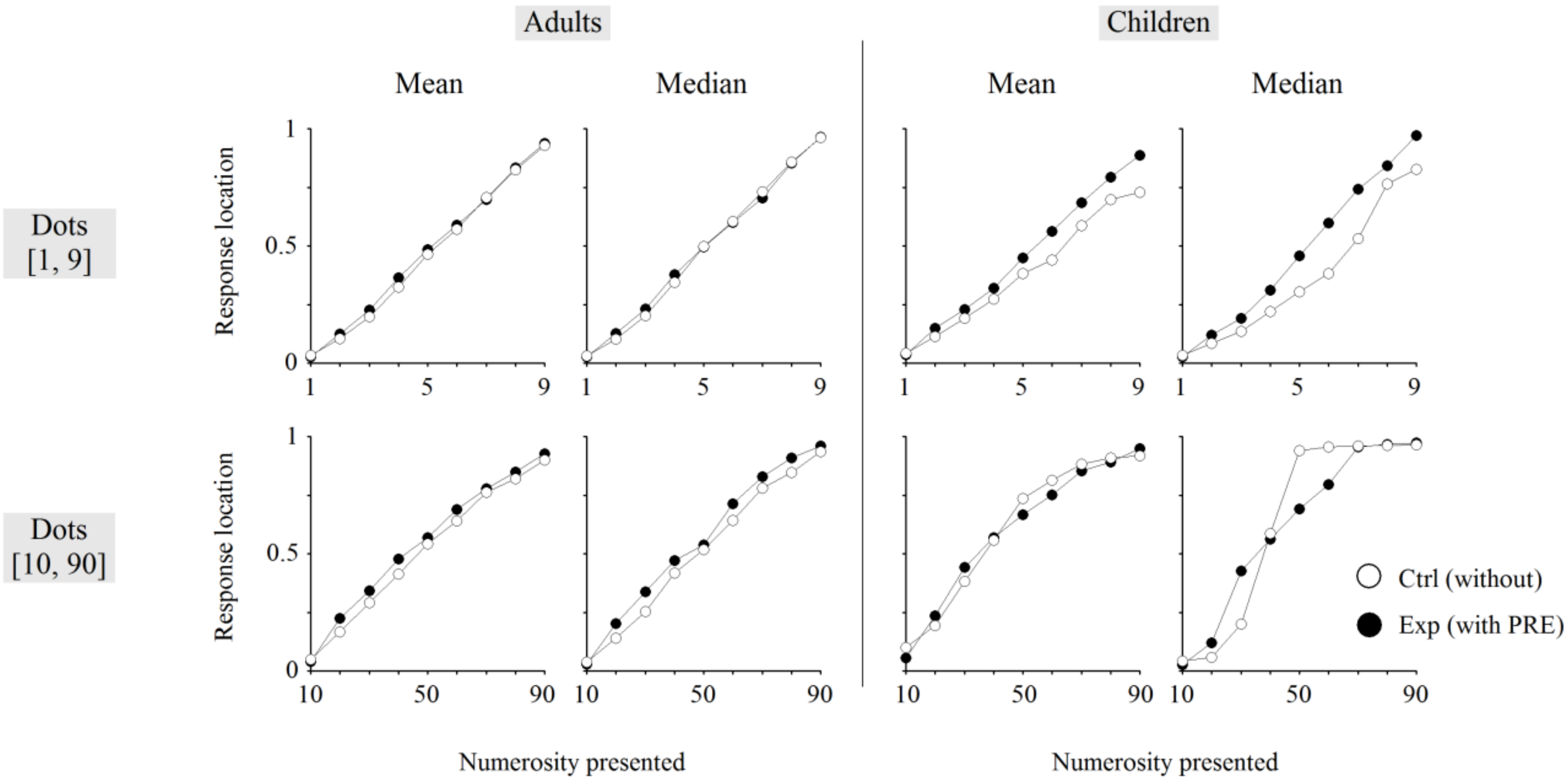


Figure 5. Numerical mapping of children and adults, for each stimulus range and experimental group. For each age group, the y-axis of the graphs on the left column depicts the mean location of responses, from all trials taken together. The graph in the right column depicts the median of these trials. Within each graph, the filled dots depict responses from the Experimental Group, which had previously experienced the brightness pre-training, and the white dots the data from the Control group.

Similarly to previous studies (e.g., Dehaene et al., 2008; Núñez, Doan & Nikoulina, 2011), OLS multiple regression analyses were carried out, contrasting the contribution of a non-linear component (β_{log}) over and above a linear regressor (β_{lin}) to judge whether a logarithmic model, rather than a linear model, would prove to be a better descriptor of the spatial mapping. These analyses confirm that a linear function best describes adults' mapping in the Dots[1,9] condition, both when one attends to the mean or to the median response location (β_{lin} significantly larger than the β_{log} regressors; see Appendix C for the full results of the multiple regression analyses).

Visual inspection of adults' group scatterplots of the Dots[10,90] range, depicted in the lower row of Figure 5, suggests a slightly more curved mapping. Whether the participants had been previously trained with brightness stimuli or not, did not seem to affect the performance, because the two group curves superimpose, be it the mean or the median response location. In addition, comparing the two central tendency measures shows no striking displacements differences. A larger displacement between group curves does seem to occur in the Median graph, with the Control group curve appearing less curved than the Experimental group. The apparent slight compression of the larger numbers at the right portion of the response bar, suggested by visual inspection, is confirmed by the OLS multiple regression analyses (Appendix C). It was found that both linear and logarithmic regressors are significant, but the logarithmic ones yielded larger weights. The only exception, as suggested by visual inspection, was the median curve from the Control group, a case where only the linear regressor is significant ($\beta_{lin} = 0.010$, $p < .001$; $\beta_{log} = 0.111$, $p = .308$).

Overall, our adults' group data agree with current literature. Similarly to American adults from Dehaene et al. (2008)'s and Núñez et al. (2011)'s data, a magnitude effect is verified, attending that smaller numbers were mapped linearly, whereas larger numerosities were distributed logarithmically along the response bar. This magnitude effect was not present in our first experiment (reported in Study 1) and in Anobile et al. (2012)'s study, in which adults mapped large sets of dots in a linear fashion. As has been discussed in Study 1, procedural differences do not seem to account for the magnitude effect. A possibility is that an uncontrolled variable such as verbal enumeration/estimation occurring concomitantly with the spatial response may have exerted differentiating effects in adults' spatial estimation. To our knowledge, no number-to-position study has ever controlled or manipulated this possible confound, and we would tentatively suggest as a first strategy the introduction of an articulatory suppression task concurrently with the presentation of the numerical stimuli (e.g., Exp. 3 in Roitman, Brannon, Andrews, & Platt, 2007; Rattat & Droit-Volet, 2012). A final comment is that attending to the mean or median response location of adults' group curves leads to identical conclusions. Hence, regarding adults' group data, our concerns with the central tendency measure seemed unnecessary.

Children. In the Dots[1,9] condition, while all children preserved the lowest anchor, ten from the Control group and three from the Experimental did not keep the largest anchor. In the right portion of Figure 5, one observes that both the mean and median curves of the Control group do not reach the right endpoint of the bar at the larger anchor. Average location for the 9-anchor in the Control group was .73 (SD = .15) or .75 (SD = .20), for the mean and median measures, respectively. As for the Experimental group curve, the 9-anchor mapping reaches up to a mean location of .89 (SD = .13) or a median of .92 (SD = .12). Independent samples t-test confirmed that performance for the largest anchor significantly differed between groups (for mean location, t(1,21.69) = 2.853, p = .009; for the median, t(1,17.8) = 2.59, p = .019).

The Experimental Group curves, with more preserved anchors, resemble adults' linear ones in the same condition. The Control Group curve behaves in a distinct manner in that, not only is the largest anchor mapping lost, but also there doesn't seem to be a constant spacing between subsequent numerosities, first with the mean curve somewhat flattening at the last three points and secondly, in the median curve, with the increasing spacing between numerosities 6 and 8. The curves in the Control group are less alike to the linear, almost diagonal ones found in the graphs of the Experimental group. Accordingly, the slope (m) of a line (y = m.x + b) fitted to the Experimental Group's mapping of the Dots[1,9] condition (Mean: m = .108, R^2 = 1; Median: m = .123, R^2 = 1) was higher than the Control Group's (Mean: m = .091, R^2 = .99; Median: m = .103, R^2 = .98).

Nevertheless, the OLS multiple regression analyses revealed that a linear function is the best descriptor of both the Experimental and Control group curves (only the β_{lin} regressor was significant for the mean location; for the median, both β_{lin} and β_{log} were significant, but β_{lin} was larger) (Appendix C).

Visual inspection of children's group scatterplots on the Dots[10,90] condition, shows that irrespective of experimental group and central tendency measure, the previously trained anchors were preserved. However, five children from the Control group and one from the Experimental group lost the smaller anchor during testing ("10-leftmost"), and five children from the Control group and the same one from the Experimental group (b10) lost the larger anchor ("90-rightmost"). Returning to Figure 5, the mean curves of the two experimental groups are similar. As numerosities increase, they tend to be mapped along the bar with progressively less spacing between subsequent samples, so that the larger numerosities are more compressed at the right side of the response bar. When one attends to the median graph, it is striking how the median curve from the Control group differs from the mean curve and from the median curve of the Experimental group. Whereas the median curve from the Experimental group is curved with larger numbers compressed at the right of the bar, the median curve from the Control group is S-shaped, suggesting clusters of responses (categorical responding). In the median Control group graph, responses are concentrated at the leftmost part of the bar for the numerosities 10 to 30, sets of 40 dots are placed slightly further to the right of the response bar's midpoint, and the remaining numerosities (50 to 90) are mapped practically at the same location, in the rightmost position of the bar (the previously trained largest anchor mapping).

The mean curves from both experimental groups were better described by the logarithmic model (β_{lin} and β_{log} with $p < .05$, and $\beta_{log} > \beta_{lin}$). The logarithmic was the only model that described the median curve from the Experimental group ($\beta_{lin} = 0.005$, $p = .193$; $\beta_{log} = 0.193$, $p < .05$). However, the S-shaped median curve from the Control group was neither described by the logarithmic nor the linear model (β_{lin} and β_{log} with $p > .05$).

If we were to analyze only mean group curves, we would have found no surprising patterns: the group curve visually approaches a straight line in the smaller range of numerosities (Dots[1,9]), and a more arched curve suggests compression of larger numerosities (Dots[10,90]). This replicates findings in literature concerning the magnitude effect, and brings the novelty of expanding the phenomenon to preschooler's nonsymbolic stimulus modality (i.e., sets of dots).

Another conclusion would be that the pre-training in a continuous non-numerical dimension had no remarkable effect. Like the Control group, the mean curve of the Experimental group for numerosities 1 to 9 is better described by a linear function, and the effect of the pre-training seemed to be the maintenance of the largest anchor. In addition, both the mean curves in the Dots[10,90] numerosities are more similar to a logarithm than a linear model.

However, these conclusions on children's performance are greatly challenged by the analysis of the median group data in the Dots[10,90] condition. This S-shaped curve, markedly divergent from the mean curve, was neither described by the linear nor the logarithmic model. Further inspection of the phenomenon is needed for the following reasons. First, it is evidence that the averaging measure one attends to may drastically change the conclusions. Secondly, because it is so different from the median of the Experimental group, it suggests that the experimental manipulation of the pre-training must have worked in the direction of different individual patterns occurring, dependent on the experimental group.

A problem with this level of analysis is that it does not inform us about the features of the individual patterns, although it strongly suggests that many children, at least in the Control group, have responded in a non-continuous fashion. That is, instead of numerosities being placed along the full extension of the bar, with locations further to the right following larger numerosities, children may have restricted their responses to some segments of the bar. In fact, this sigmoid type of responding is reminiscent of the psychometric curves derived from the two-alternative forced-choice (2AFC) methods. In other words, if many children were to respond by alternating between the two anchor positions, than they would be in a 2AFC task, such as in the numerical bisection task.

We will thus advance with the analysis at progressively deeper units of analysis: inspection of individual averages and of trial-by-trial responses, whilst commenting on the resolution to detect and screen for possible non-continuous patterns.

3.2.2. Individual analysis - simple curve fitting

Ever since the seminal work by Siegler and Opfer (2003), a usual metric for the comparison between the logarithmic and the linear model is to contrast the variance accounted for by each model. The authors describe each participant's performance according to the model with higher R^2 values, and consequently code participants as "Logarithmic", "Linear" or "None" (if neither fitting reaches statistical significance). Lastly, they compare the percentage of participants on each case, contrasting smaller with larger range conditions.

As such, in the current study average response location of each participant was fitted to simple linear (y = m.x + b) and logarithmic ($y = m.\log_{10}(x) + b$) functions. As it happened at the group level analysis, both mean and median response locations were inspected, to investigate whether they lead to similar conclusions regarding the log vs. linear comparison (Tables 1 and 2, respectively).

Adults. In the Dots[1,9] condition all participants, from the Control and Experimental groups, present higher coefficients of determination (R^2) in the fitting of the linear model to their mean response location, than the logarithmic model (refer to Table 1 for the complete individual scores). As for the larger, Dots[10,90] condition, 6 out of the 24 participants' mappings were better fitted by a logarithmic than a linear function (m11, w15, and m21 in the Control Group; m4, m9, and w17 in the Experimental group). The remaining (75%) of the participants' mappings were better fitted by a linear function.

Applied to median response location, the log vs. lin comparison reached the same overall conclusions. First, in the Dots[1,9] condition all adults' median response locations were better described by a linear model (refer to Table 2 for the complete individual scores). Second, in the Dots[10,90] condition, 7 out of the 24 participants' mappings were better fitted by a logarithmic than a linear function (the same six previously mentioned in the mean analysis plus participant w24 from the Experimental group). The remaining (71%) of the participants' mappings were better fitted by a linear function.

Table 1

R^2 values yielded by the fitting of simple linear and logarithmic functions to each participant's <u>*mean*</u> *response location and to the group mean curve.*

	Children					Adults				
		R^2					R^2			
		Dots [1,9]		Dots [10,90]			Dots [1,9]		Dots [10,90]	
		Lin	Log	Lin	Log		Lin	Log	Lin	Log
Ctrl.	**b19**	.89	.91	.95	.82	**w15**	.97	.81	.95	.98
	g8	.92	.74	.97	.87	**m21**	.99	.89	.89	.97
	g15	.99	.87	.84	.94	**w16**	.97	.80	1	.90
	g16	.88	.98	.81	.90	**m11**	.99	.94	.94	.99
	b22	.74	.51	.85	.90	**m20**	1	.90	.98	.94
	b17	.82	.87	.93	.86	**w14**	.98	.84	.97	.93
	g13	.95	.92	.83	.88	**w13**	1	.90	.97	.82
	b4	.93	.75	.86	.96	**m22**	.99	.88	.92	.91
	g14	.96	.92	.86	.84	**w12**	1	.90	.98	.89
	g5	.99	.89	.82	.95	**m18**	.99	.89	.99	.87
	b18	.83	.62	.93	.94	**m19**	.96	.77	.97	.85
	b9	.93	.75	.69	.87	**w23**	.99	.91	.98	.92
	Avg	.90	.81	.86	.90	**Avg**	.99	.87	.96	.92
	SD	.07	.14	.08	.05	**SD**	.01	.05	.03	.05
	Group	.99	.88	.92	.95	**Group**	1	.88	.99	.94
Exp.	**b23**	.97	.81	.92	.99	**w17**	.99	.89	.88	.98
	g20	.98	.93	.96	.95	**m5**	1	.88	.98	.95
	g11	.82	.93	.71	.87	**w1**	.98	.94	.98	.91
	b2	.98	.91	.99	.91	**m4**	1	.90	.88	.99
	g1	.90	.69	.91	.97	**w24**	.99	.93	.98	.97
	g7	.99	.88	.91	.99	**m3**	.99	.89	.95	.91
	b10	.74	.90	.96	.96	**w6**	1	.92	.97	.97
	b3	.97	.85	.75	.91	**w10**	.99	.86	.97	.96
	g6	.92	.76	.93	.77	**m7**	1	.91	.99	.92
	b12	.99	.88	.73	.93	**m9**	1	.91	.95	.98
	b24	.94	.75	.95	.98	**m2**	.99	.87	.97	.93
	g21	1	.90	.99	.93	**w8**	1	.89	.95	.79
	Avg	.93	.85	.89	.93	**Avg**	.99	.90	.95	.94
	SD	.08	.08	.10	.06	**SD**	.00	.02	.04	.05
	Group	1	.89	.95	.99	**Group**	1	.90	.98	.97

Table 2

R^2 values yielded by the fitting of simple linear and logarithmic functions to each participant's median response location and to the group median curve.

	Children					Adults				
		R^2					R^2			
		Dots [1,9]		Dots [10,90]			Dots [1,9]		Dots [10,90]	
		Lin	Log	Lin	Log		Lin	Log	Lin	Log
Ctrl.	**b19**	.77	.71	.81	.61	**w15**	.97	.81	.96	.98
	g8	.91	.71	.84	.71	**m21**	.99	.89	.88	.96
	g15	.96	.83	.81	.90	**w16**	.97	.80	1.0	.89
	g16	.83	.91	.68	.74	**m11**	.99	.95	.95	.98
	b22	.53	.32	.70	.81	**m20**	1.0	.91	.98	.94
	b17	.72	.75	.77	.70	**w14**	.98	.83	.96	.91
	g13	.97	.94	.69	.75	**w13**	1.0	.90	.96	.81
	b4	.94	.80	.80	.92	**m22**	.99	.88	.90	.90
	g14	.95	.92	.76	.72	**w12**	1.0	.90	.99	.88
	g5	.95	.92	.81	.94	**m18**	.99	.89	.98	.86
	b18	.79	.58	.90	.89	**m19**	.93	.75	.93	.83
	b9	.86	.68	.65	.81	**w23**	.99	.91	.97	.90
	Avg	.85	.76	.77	.79	**Avg**	.98	.87	.95	.90
	SD	.13	.18	.08	.10	**SD**	.02	.06	.03	.05
	Group	.95	.77	.82	.86	**Group**	1.0	.89	.99	.92
Exp.	**b23**	.96	.81	.88	.96	**w17**	.99	.89	.88	.98
	g20	1.0	.91	.96	.93	**m5**	.99	.88	.97	.95
	g11	.83	.82	.54	.73	**w1**	1.0	.93	.97	.88
	b2	.96	.90	.96	.81	**m4**	.99	.90	.89	.99
	g1	.87	.67	.85	.87	**w24**	.99	.92	.97	.98
	g7	.98	.85	.88	.95	**m3**	.99	.89	.96	.91
	b10	.64	.83	.96	.96	**w6**	1.0	.92	.97	.95
	b3	.97	.84	.66	.83	**w10**	.99	.85	.96	.94
	g6	.91	.76	.83	.65	**m7**	.99	.92	.96	.88
	b12	.99	.88	.68	.89	**m9**	1.0	.91	.96	.97
	b24	.91	.74	.94	.96	**m2**	.98	.84	.97	.93
	g21	1.0	.91	.99	.93	**w8**	1.0	.89	.93	.76
	Avg	.92	.83	.84	.87	**Avg**	.99	.90	.95	.93
	SD	.10	.08	.14	.10	**SD**	.00	.03	.03	.06
	Group	.99	.87	.93	.96	**Group**	1.0	.90	.99	.95

Children. In the Dots[1,9] condition, three children's mean mappings in the Control group were better described by the logarithmic model (g16, b17, and b19). In the Experimental group, two were also classified as logarithmic (b10, and g11). The remaining participants' mappings (79%) were better described by the linear model. In the Dots[10,90] condition the pattern reversed: four children in the Control group (g8, g14, b17, and b19), and in the Experimental group (b2, g6, g20, and g21) were classified as linear, and the reaming 67% as logarithmic.

As for the median data, whereas the data of child b22 was not significantly fitted by the logarithmic model ($p = .11$), the remaining children were significantly fitted both by the linear and the logarithmic model ($p < .05$). Analysis of median response location confirms the pattern found with the mean measure. In the Dots[1,9] condition, the same three children in the Control group (g16, b17, and b19) and one from the Experimental group (b10) are better classified as logarithmic. The remaining 83% of the participants' mappings were classified as linear. In the larger Dots[10,90] condition, five children from the Control group (g8, g14, b17, b19, and b18) and five from the Experimental Group (b2, g6, g20, g21, and b10) were classified as linear, and the remaining 58% as logarithmic.

Were we to stop here and we could be satisfied to assume that the group analysis summarizes adequately individual behavior. On one hand, the fitting of children's individual average curves was consistent with the group level analysis (Figure 5). To recap: a preferably linear pattern occurs with smaller numerosities (Dots[1,9]) but, for larger numerosities (Dots[10,90]) the performance is best described by the logarithmic model. Additionally, in the current study no individual curve fitting, in children or adults, failed to reach significance for either of the two models. As for the adults' data, while the group curve of the Dots[10,90] condition was better fitted by the logarithmic model, the majority of the individual curves were better fitted by a linear function. Nonetheless, the frequency of the "Logarithmic" cases did increase in the larger Dots[10,90] range.

We present three final comments on the goodness of fit of linear and logarithmic models to individual average response locations. First, participants have to be dichotomously classified as one of the models, but may present near R^2 values on both fittings. Sometimes the difference between the R^2 is very small (e.g., the differences for b10 and w6 are on the order of the thousandth, see Table 1). Second, significant and relatively high R^2 values may result from the fitting, but the average curve behaves in a distinct manner from the fitted model. The fact that high R^2 may be obtained by a model that poorly characterizes the data is a known precaution in statistical fitting procedures (e.g., Roberts & Pashler, 2000; Taatgen & Van Rijn, 2010). Nonetheless, the debate on how worthwhile are R^2 comparisons has been absent in number-to-position studies. In the current study, particularly in the Dots[10,90] condition, many child participants presented average curves that were S-shaped. Recall that the median response location of the Control group, at the group level (Figure 5) was also S-shaped and suggested the occurrence of non-continuous, categorical responding. Categorical responding would also entail that, in a within-participant comparison between the mean and median measures, the latter would reveal steeper sigmoid-like curves, which was the case with our results (refer to Appendix D for the complete individual plots of mean and median average location). Lastly, the level of analysis that found significantly good fitting of linear and logarithmic models to individual averages was unable to detect non-continuous cases. As we will next show, only at a finer level – the inspection of single trial responses – will it be possible to reveal children's patterns and, in addition, to expose the disparity when interpreting children's performance across the different units of analysis.

3.2.3. Individual analysis – single-trial scatterplots

Following the review of the scarcely reported individual performance as well as the evidence from our previous study (Study 1), the main interest of this study was to analyze individual performance at the trial level and to verify the consistency between results taken at this and the supra-, group unit of analysis.

Adults. To better illustrate the overall individual patterns, we selected some exemplars (refer to Appendix E for the scatterplots of the 24 adults). Figure 6 shows the performance of six subjects, three from each experimental group. These exemplars were selected to illustrate the different types of individual responding. As depicted in the first row of the graph, in the Dots[1,9] condition adults from both groups placed the numerosities almost equally spaced along the response bar. All preserved the 1-leftmost anchor that they had been trained on. Only three participants did not keep the 9-rightmost anchor, responding at about the .78 location of the bar (m19, m2 and w23;

the last two depicted in Figure 6). Notably, these three cases were those who mostly presented an increase of response variability (noise) with the numerosity. Nonetheless, for the most part adults produced high precision across all numerical samples, particularly the anchors.

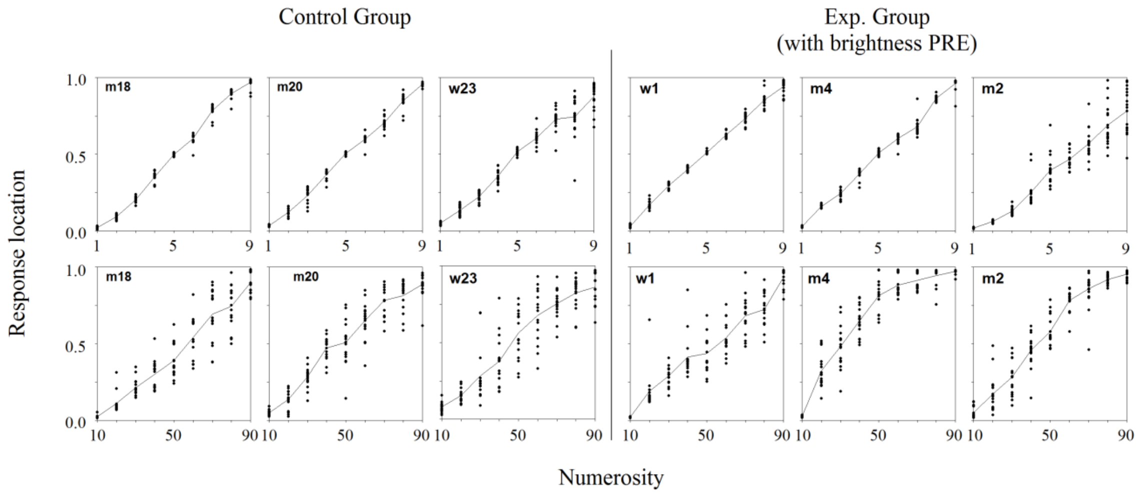


Figure 6. Results of six adult participants, m18, m20 and w23 from the Control Group, and w1, m4 and m2 from the Experimental Group. The first row depicts the Dots[1,9] condition, and the second row the Dots[10,90] one. Each data point is the response location on a single trial; the line is the mean location.

Inspection of the second row of the graph, regarding the Dots[10,90] condition, immediately suggests that the data points are more spread (i.e., less precise), in comparison with the smaller range condition. This decrement in response precision with larger numerosities was found in all participants regardless of their experimental group. One can observe that both the “spread” of the dots’ and the individual mean curves, rise with the presented numerosity. In other words, with increasing samples, participants responded progressively towards the right side of the bar. Consequently, the individual average curves are representative of adults’ continuous pattern of responses along the bar.

To quantify the use of the bar one may recur to the normalized entropy measure (H). For each participant’s numerical session, we computed a 9 bin histogram of all response locations. Next, H scores were estimated according to the formula $H = -\frac{\sum_{i=1}^{9}(p_i \times log_2 p_i)}{log_2 9}$, where p_i is the relative frequency of responses on each bin.

As visual inspection of their scatterplots has shown, all adult participants responded evenly alongside the full extent of the bar. This is illustrated by their high H scores, close to the possible maximum (H = 1) (Figure 7). In the Dots[1,9] condition, H scores averaged .96 (SD = .04) in the Control group and .97 (SD = .02) in the Experimental group. These values slightly decreased in the Dots[10,90] condition, with an average of .93 (SD = .05) in the Control group and of .93 (SD = .04) in the Experimental group.

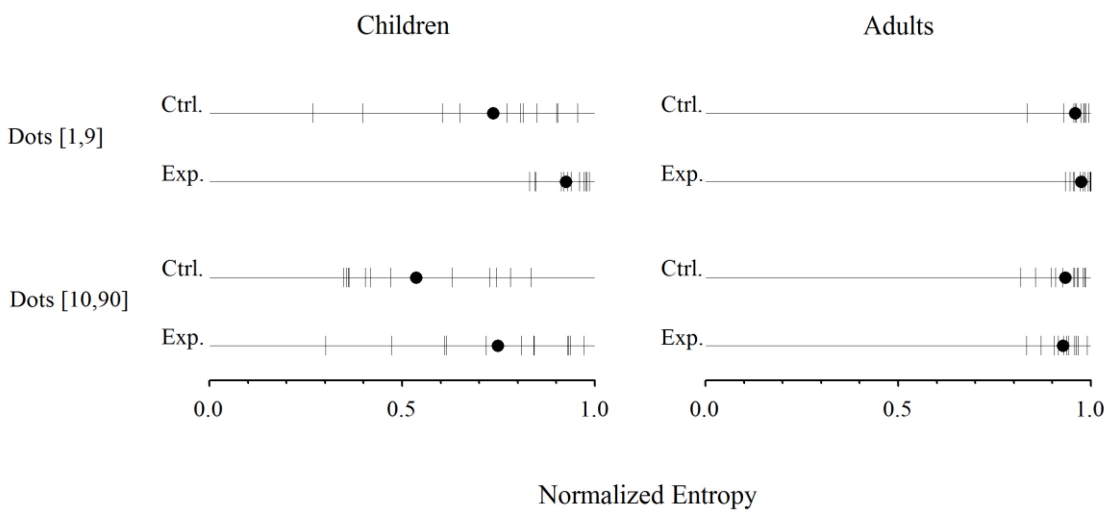


Figure 7. Individual normalized entropy scores (H). In each numerical range, the H scores are separated by experimental group (thus, n = 12 on each graph). A vertical line represents a participant's H score, and the black dot the group's mean.

A mixed between-within subjects ANOVA was conducted to compare H scores between participants in the Experimental and Control groups, across the two tested numerical ranges (Dots[1,9] and Dots[10,90]). There was no significant difference in H scores between the two groups ($F(1,22) = 0.114$, $p = .739$). Additionally, no significant interaction was found between group and range, $F(1,22) = 1.662$, $p = .211$). There was, however, a significant main effect for range ($F(1,22)= 16,753$, $p < .001$). As illustrated in the right portion of Figure 7, compared to the

Dots[1,9] condition, entropy scores in the Dots[10,90] condition were significantly lower, both for the Control and the Experimental group (refer to Appendix F for a table with the complete individual H scores).

Thus, analysis of the H scores and visual inspection of individual-trial scatter plots confirm that (i) adults' average response location is a good indicator of their individual performance, since all adults responded in a continuous pattern; (ii) the experimental manipulation of the pre-training in brightness did not lead to differences between the groups regarding the use of the response bar; and (iii) the comparatively lower utilization of the response bar in the larger numerosity condition did not reflect the occurrence of a non-continuous pattern of responses, nor a preference for a logarithmic-like mapping, but merely the occurrence of "gaps" or jumps between successive numerosities. In other words, contrary to the Dots[1,9] condition, where responses are highly precise and occur at equally spaced intervals, in the Dots[10,90] condition some individual scatterplots present empty intervals. To illustrate, in participant m4's graph of Figure 6, the initial portion of the bar was less utilized. If we project the data points onto the y-axis, there would be a "gap" without any responses between the smallest anchor-10 and the next lower numerosity, 20. Additionally, lower H scores did not necessarily result from a more logarithmic-like pattern. Only three out of the six participants, whose mean response location was better described by the logarithmic than the linear model (R^2), are on the bottom one-third of the H score ranking.

Children. As with the adults' data, we selected three cases of each experimental group to illustrate the overall patterns found in children's trial scatterplots. Visual inspection of the complete scatterplots led us to identify three main response patterns. Before advancing with their description, it is worth recalling that all participants' average location increased as a function of numerosity (refer to Appendix E for the scatterplots of the 24 children). First, some participants restricted their responses to the anchor positions. In Figure 8, participant b19 exemplifies this bi-categorical pattern, both in the Dots[1,9] and the Dots[10,90] conditions. As can be observed, during testing the participant distributed his responses amongst the two (previously trained) endpoints of the bar, in differentiated proportions so that mean response location increased with numerosity. Recall that this is the type of categorical responding ("virtual" 2AFC) which had previously been hinted at when considering the other units of analysis (group to a certain extent, but mostly individual average curves).

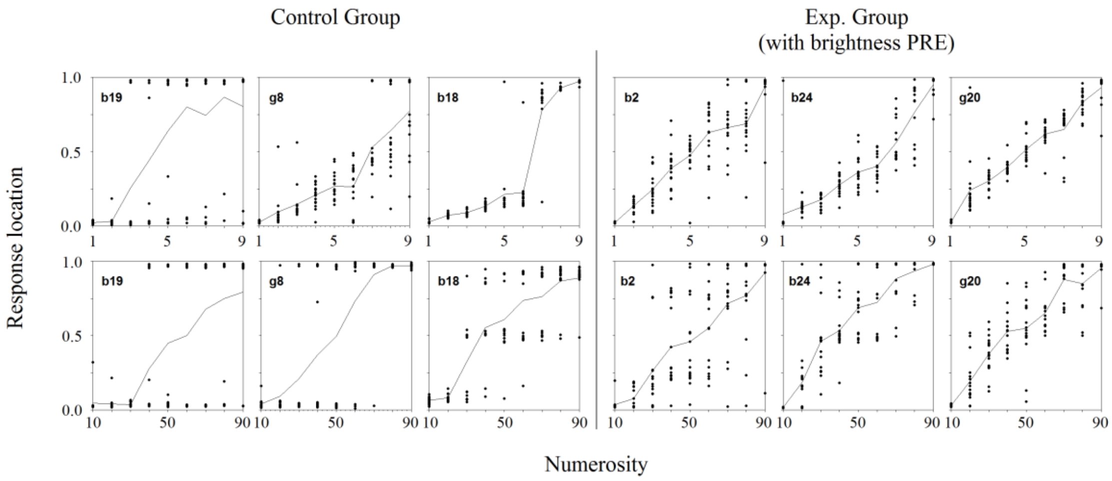


Figure 8. Results of six child participants, b19, g8 and b18 from the Control Group, and b2, b24 and g20 from the Experimental Group. The first row depicts the Dots[1,9] condition, the second row depicts the Dots[10,90]. Each data point is the response location on a single trial; the line is the mean location.

Secondly, other participants responded at the anchors plus a third location on the bar, which was near the midpoint – tri-categorical pattern. This is exemplified in Figure 8 by the performance of b18 in the Dots[10,90] condition. Lastly, there were also children who presented a broader distribution of their responses along the bar. To exemplify, in Figure 8 the number of response clusters surpasses three for participants g8 and b18 in the Dots[1,9] condition, and for b2, b24 and g20 in both range conditions.

We determined whether the number of clusters that best describe each child's distribution of responses along the bar was 2, 3 or more. This cataloging was undertaken by visually inspecting the individual trial scatterplots and by performing a k-means cluster analysis to determine the cut-off point for the number of utilized portions of the bar. In Figure 9, two histograms represent the percentage of these patterns, separated by range condition and experimental group.

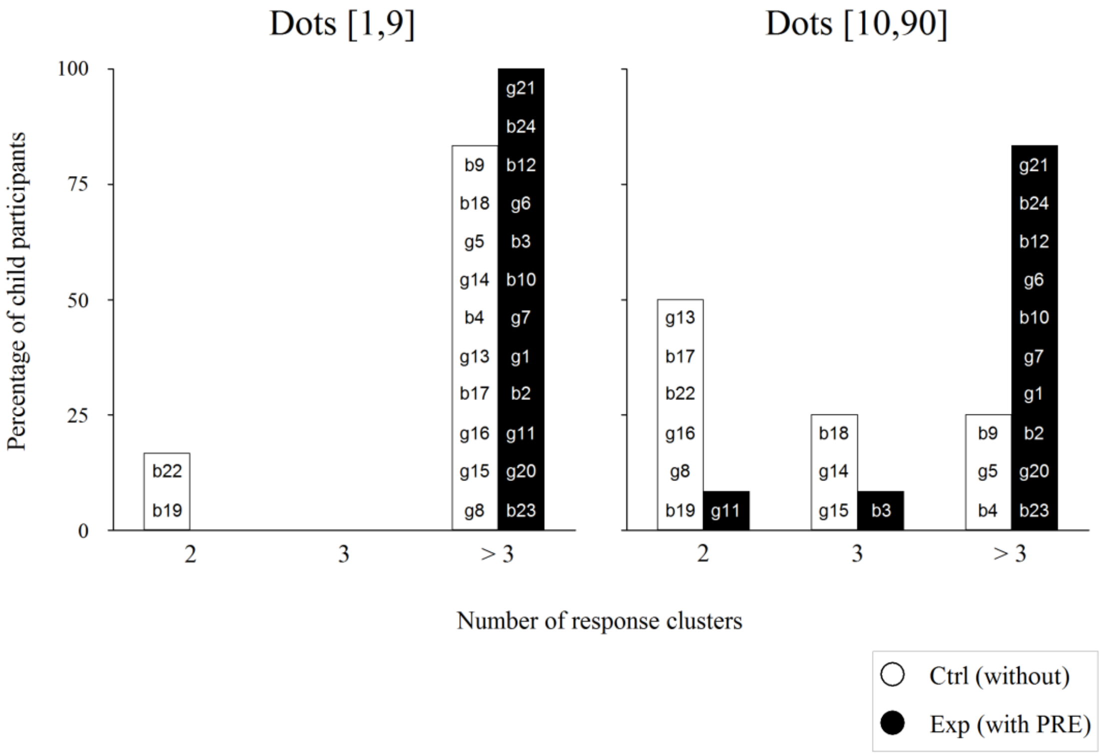


Figure 9. Percentage of children whose response distributions are better described as two, three or more response clusters.

In the Dots[1,9] condition, except for two participants in the Control group, children responded at more than three regions of the bar. It is, nevertheless, difficult to argue in favor of them presenting a continuous pattern similar to the continuous pattern emitted by adults (independent samples t-test on H scores revealed a significant difference between children's (M = .83, SD = .18) and adults' (M = .97, SD = .01) use of the response bar; $t(2,46) = -3.61$, $p < .001$). There are cases, as exemplified by participant b18 in Figure 8, where response location seems to increase linearly from numerosity one to five or six and remain constant afterwards, at the rightmost location of the bar. Yet another divergence from adults' performance is that there are cases where the mean and "spread" of response locations increases linearly with numerosity, with responses moving towards the right side of the bar, but with maximum response location peaking quite behind the rightmost endpoint of the bar. This occurred predominantly in the Control group children. Related to this, and as mentioned above, is the fact that more children in the Control group lost the 9-rightmost anchor mapping (10 in the Control group, opposed to 3 in the Experimental group).

Accordingly, the inferential statistics revealed a significant main effect of experimental group on the entropy scores ($F(1,22) = 10.635$, $p < .01$), which suggests that the children who had experienced the brightness-to-position pre-training tended to distribute their responses more broadly along the bar ($M = .92$, $SD = .06$), compared to the children in the Control group, solely tested in the number-to-position tasks ($M = .74$, $SD = .22$) (see Figure 7 and Appendix F).

The number of participants who restricted their responses to the endpoints or to the endpoints plus the middle of the response bar increased more than fivefold in the Dots[10,90] condition. The analysis of the entropy scores is consistent with the reduced use of the bar with larger numerosities, revealing a main effect of range, $F(1,22) = 21.064$, $p < .001$. In fact, 75% of the participants in the Control group were classified as bi- or tri-categorical. Not surprisingly, the entropy scores were significantly lower in the Control group ($M = .54$, $SD = .19$) than in the Experimental group ($M = .75$, $SD = .21$), and there was no interaction effect between range and experimental group ($F(1,22) = .084$, $p = .774$) (see Figure 7). The smaller entropy scores tended to belong to the participants whose responses were concentrated in two or three response clusters. Indeed, the Spearman's rank correlation between entropy scores and number of response clusters was highly significant (ρ (23) = .85, $p < .001$). As for the ">3" cluster patterns, the Experimental Group had more than thrice of these cases than the Control Group. There is yet another observation pertaining to the ">3 cluster" patterns emitted with larger numerosities. Consider the performance of participants b2 and b24 during the Dots[10,90] condition in Figure 8. Certainly, there are more than three response clusters. However, they are not necessarily a continuous pattern, at least not in the same sense as adults' participants (Figure 6) (independent samples t-test on H scores revealed a significant difference between children's ($M = .64$, $SD = .22$) and adults' ($M = .93$, $SD = .05$) use of the response bar; $t(2,46) = -6.14$, $p < .001$). Contrary to the rightward moving spread of responses along the bar observed, for example with g20, in these participants' scatterplots the average line befalls in-between successive clusters. Thus, our cluster-based categorization does not fully capture all the idiosyncrasies of continuous response patterns. Nonetheless, it was helpful to measure the effect of the experimental manipulation of experiencing a pre-training in brightness, prior to the testing with numerical stimuli.

3.3 Counting and Verbal Estimation Assessment

3.3.1 Abstract counting

Six children did not enter the Counting and Verbal Estimation Assessment phase because they left the kindergarten earlier than the rest of their classmates. In the first task of this phase, the experimenter asked the child up until what number he/she could count and, in order to validate this self-reported knowledge, the child was subsequently tested with decade transition sentences (e.g., continue after me "56, 57,"…). The median self-reported number was "29", with the range from "4" (b22) to "200" (b18). The self-reported measure was, for the most part, congruent with the decade transition performance, given that the overall median difference between these scores was 0. The discrepancies were the two children who had overestimated their counting ability (by a difference of 20), and the three whose self-report underestimated their ability (notably, child b18 counted up until 349, at which point the experimenter told him he could stop). In these cases, the decade transition score overruled the self-reported one. Except for b22 who merely recited the sequence of words up until 4, all children counted at least up to 10. Only 26% recited the number words sequence for more than 30.

Previous studies that have tasked preschoolers with both number-to-line tasks and counting assessment usually opted to introduce a cut-off point and divided children into groups of skilled vs. unskilled counters. Next, they carried out group comparisons on the slope and goodness-of-fit of the linear regression between numerical sample and response location (e.g., Ebersbach et al., 2008). As we have previously discussed, our results discouraged us from applying this comparison. Therefore, the abstract counting score will solely be taken for the correlations with other counting and number-to-line measures (Soltész, Szucs, & Szucs , 2010), which we will address further in this text.

3.3.2 Counting Objects

The Counting Objects tasks started with the child being presented nine horizontally aligned marbles, and being asked to count them aloud, one-by-one. All children counted the nine marbles correctly. Even child b22, who in abstract counting had been able to recite only the number words from one to four, performed accurately.

Briars and Siegler (1984) conceptualize learning to count as an induction task, in the sense that while watching the counting of others, such as teachers and parents, "children must induce which features are essential for correct counting and which are optional". And though the only essential feature is that each item is assigned merely one number word (the "one-to-one" principle from Gelman and Gallistel (1978)'s counting criteria), four optional features are typically present during this learning: (a) the subject counts adjacent objects consecutively, (b) the subject points once to each object, (c) the subject starts at an end of a row, rather than its middle, and (d) counting proceeds in a left to right direction (Briars & Siegler, 1984). All of our participants showed features (a), (b) and (c) and only three children (g15, g20, and b23) did not show the left-to-right directional bias (d).

This directional bias is usually found in Western children and adults (Dehaene, 1997, p. 82) and has also been reported by Opfer and Thompson (2006) and, later on, by Opfer, Thompson, and Furlong (2010), the authors from whom we adapted the counting tasks. In this latter study (2010), 76% of all the tested 2.5- to 8.4-year old children, and 100% of the adults presented a left-to-right directional bias. Dehaene (1997) suggests that the left-beginning to right-endpoint of counting is most likely a consequence of the Western writing system. Besides being learned during school, it is ubiquitously present in the environment, in "rulers, calendars, mathematical diagrams, library bookshelves, floor signals above elevator doors, computer keyboards, and so on" (1997, p.82; Dehaene, Bossini, & Giraux, 1993, Experiment 7).

The three remaining counting tasks also related to the relationship between number of marbles and spoken numerals, from 1 to 9 ("give me {numeral} marbles.", "how many marbles are there?", and "give me the same number of marbles."). Figure 10 depicts the results from these three tasks, collapsed across all trials. Children performed accurately in all tasks, as indicated by all colored series peaking at the correct numerosity, with the scarce errors occurring at neighboring values. The frequency of errors tended to increase with the value of the numerical target. Errors mostly consisted of the subjects overestimating the required numerosity, that is, emitting a response larger than the target numerosity. In fact, out of the total errors the percentage of overestimates in the "Numerals to Marbles", "Marbles to Numerals" and "Marbles to Marbles" tasks was 72%, 73% and 75%, respectively.

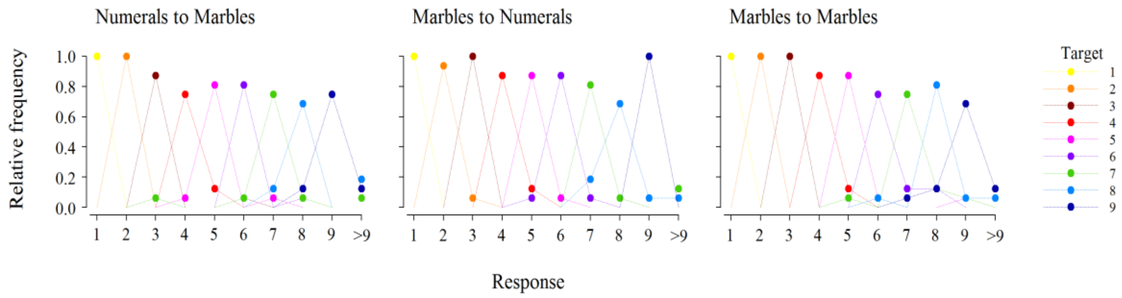


Figure 10. Results from the three Counting Objects tasks. Subjects were mostly accurate, as seen by responses peaking at the target (i.e., asked) numerosity. In the first and third graphs, the participant's response was to give a certain number of marbles. In the second graph, the participant had to enumerate sets of marbles.

Children b10, g11 and b22, who performed the worst in the "Numerals to Marbles" (Give-N) task, performed also the worst in the "Marbles to Numerals" ("how many?"). The other children correctly offered the last number word in the counting series as indicating the number of present items. This proficiency is usually interpreted as evidence that the child possesses the "cardinal word principle" (Wynn, 1990, 1992). In other words, the child understands that the last number word refers to or is a property of the entire set of items (Le Corre & Carey, 2007; Lee & Sarnecka, 2010). Yet an alternative interpretation is that the children responded mechanically with the last number word in the counting series, without truly understanding cardinality (the "last word rule" by Furlong, 1988, cit. by Bermejo, 1996, 2004; Frye et al., 1989).

Each participant was scored on his/her counting objects proficiency, according to the number of correct responses, from a minimum of 0 to a maximum of 27 (9 trials x 3 tasks). A high number of children presented scores of 25 or more (68%). Indeed, the high accuracy showed by the majority of our participants does not allow us to discriminate between good and bad counters, as was the case in Opfer and Thompson (2006)'s study. Our data are actually more similar to their "Mental-Number-Line" (best counters) group than to the "No-Mental-Number-Line" (worse counters) group scatterplot (2006, Exp.2, Fig. 6).

We believe that more exploratory analyses, beyond the inspection of the R^2s resulting from group or individual linear fittings of the estimates, or of mean group reaction times, need to be carried out before one can suggest that young children's performance in these counting tasks reflects directionally-biased representations of increasing numerical magnitude. And as Lee and Sarnecka (2010) point out, even to determine the knower level (i.e., the maximum cardinal meaning that a child understands) in Give-N ("Numerals to Marbles") tasks, one has to account for task-specific influences.

Regarding possible task-specific influences, our notes indicate certain peculiarities in the patterns of responses, whose impact on overall accuracy would be interesting to explore in the future. For example, in the "Marbles to Marbles" task some children, after counting the sample set, assembled their set by adding each additional marble in a left-to-right order. In another example of a different response strategy, other children started the trial by overtly counting the marbles presented by the experimenter then proceeded to place their own marbles, one after the other, in front of the experimenter's set and copying its spatial configuration. Thus, it was possible for them to perform accurately without resorting to counting, by matching the spatial arrays of marbles. If that was the case, then it is difficult to come up with a discussion on the nature of numerical magnitude representation. As a final note, it would be interesting to know how our children would have fared had the procedure been so that the marbles were occluded before the subject could start his response (e.g., Huttenlocher, Jordan, & Levine's (1994) study with toddlers). This manipulation would probably hinder responding based on simple perceptual matching-to-sample strategies. The task's difficulty level would thus possibly increase, which in turn would have allowed for a greater differentiation of participants by level of counting proficiency.

3.3.3 Verbal Estimation

The final two experimental sessions, for both children and adults participants, required them to verbally estimate (i.e., enumerate) simultaneous arrays of dots, ranging from 1 to 9 and 10 to 90 dots. The numerical target remained on screen until a response was provided. Participants were instructed to try and give a "first-impression" on the number of dots presented, and to respond as fast as possible. The results are presented in Figure 11.

Adults. As depicted in the group histogram in Figure 11, all number words between 1 and 9 occurred in equal frequency, hence the bars at about 0.11. And as confirmed by the lines present in the scatterplot of Figure 11, these verbal estimates were accurate. In fact, out of the 864 total estimates, only 6 were incorrect (4 errors following sample “7”, and 1 error each in samples “8” and “9”), which computes as an error percentage of less than 0.7%. No adult was perceived by the experimenter as counting the dots aloud, or pointed to the individual dots. Of course, it is likely that they counted the elements sub-vocally. Interestingly, though, even in the most adverse conditions for counting, such as when the target is presented briefly, the participants are required to give highly speeded responses and/or they are hindered from sequentially foveating individual items, adults’ estimations of small numbers are typically accurate (Simon & Vaishnavi, 1996; Schleifer & Landerl, 2011).

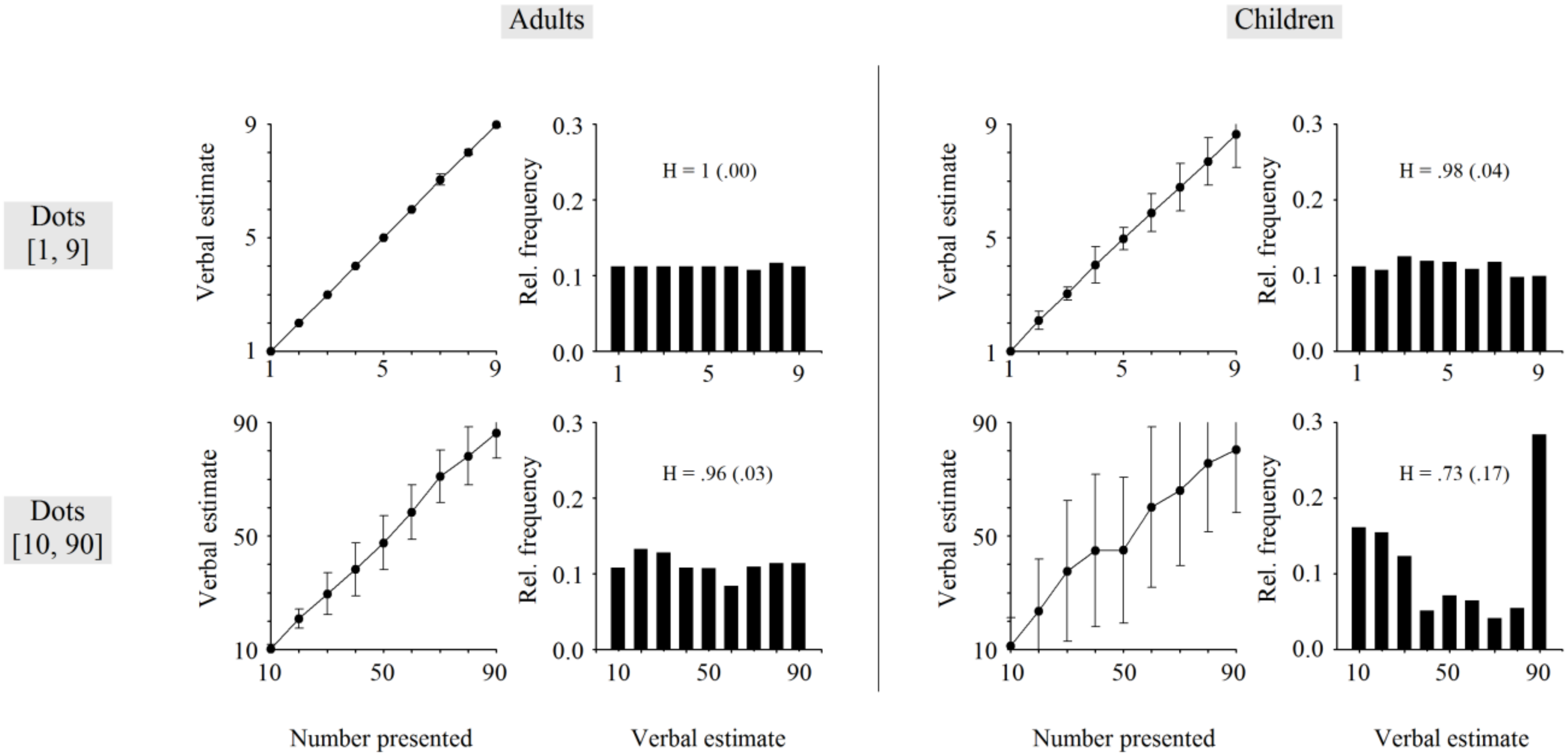


Figure 11. Results of the Verbal Estimation tasks. On each age group section, the left graphs present the group's mean (± 1 SD) verbal estimates (responses) as a function of the sample numerosities. The right graphs are the group's histogram of the verbal estimates, and also the mean and standard deviation of the individual entropy scores.

In the session with sets of 10 to 90 dots, adults' individual H scores were very high, with the verbal estimates occurring in similar frequency (the mean of the individual H scores was .96, with standard deviation of .03). Similarly to other studies investigating adults' estimation of large numbers, we observed that the estimates increased monotonically with the sample (linear regressions of individual mean estimates yielded R^2 between .91 and 1.0, mean = .97). Young adult's mean estimates have been reported to increase monotonically when they were asked to reproduce an Arabic digit in number of key-presses (sample range [7,25] in Exp. 1 from Whalen, Gallistel, & Gelman, 1999; sample range [2,32] in Cordes, Gelman, Gallistel, & Whalen, 2001), to verbally estimate how many times a figure was flashed on screen (sample range [7,25] in Exp. 2 from Whalen, Gallistel, & Gelman, 1999; sample targets {8, 16, 32} in Exp.1 and {8, 11, 14, 16, 20} in Exp. 2 and Exp. 3 from Boisvert, Abroms, & Roberts, 2003; sample range [1,20] in the "Report" condition from Tan & Grace, 2012), or to verbally estimate sets of briefly presented simultaneous elements (sample range [4,120] in Lipton & Spelke, 2005).

The analysis of the normalized entropy scores did not capture an additional feature of our adults' verbal estimation of large numbers. The overall distributions of estimates are displayed in Figure 12, showing very few colored data points projecting onto the x-axis at values in-between the decade numbers. In fact, 97.7% of the estimates were multiples of 10.

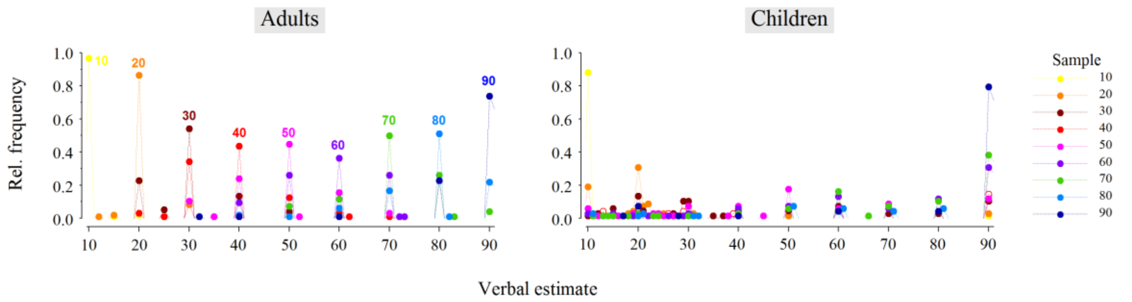


Figure 12. Distribution of the verbal estimates produced in the Dots[10,90] range. The graphs depict the produced relative frequency of each number word, and the colored lines discriminate the samples for which the particular number word was emitted.

This finding of adults' estimates occurring practically at the decade numbers throughout the sample range has also been reported by some other verbal estimation studies (Dehaene, Dupoux, & Mehler, 1990; Dehaene & Mehler, 1992; Lipton & Spelke, 2005). Adults' propensity for disproportionately producing round numbers has been suggested to derive from the frequency of these words in the language (Dehaene, 1997, pp. 82, 108), and from a possible customary strategy to offer certain numerals as "referents", for approximation (Dehaene & Mehler, 1992).

We also found that the standard deviations increased up until sample "30", at which point they remained constant across numerosities (at about 9.5). Accordingly, the coefficient of variation (CV = SD / Mean) was about .15 in samples "10" and "20", peaked at sample "30" (.25), then decreased successively until a minimum of .10 for sample "90", so that the scatterplot (not depicted) resembled an inverted U shape. A different result was reported in Whalen et al. (1999), Cordes et al. (2001), Boisvert et al. (2003) and Tan and Grace (2012)'s studies, where both mean and standard deviation increased proportionally with sample numerosity so that the coefficients of variation were constant across larger numerosities. Such estimation with scalar variability is usually assumed as an output of a mechanism that generates analog magnitude representations of numbers (e.g., Gelman & Gallistel, 1978; Gallistel & Gelman, 1992, 2000; Dehaene, 1997, 2001).

There are critical procedural features that distinguish our study from the ones reporting scalar variability of estimates, particularly the features related to sample mode and duration, as well as the information available to the participants. Namely, besides the instructions to avoid counting the sample, these studies have employed brief sample presentations (less than 1 second in duration), implemented a short response time window, did not provide information about the anchor values nor gave feedback, and/or employed a distractor task with the intent of discouraging participants from verbally counting the items. These procedural differences may explain why our adults' verbal estimates failed to show the signature of numerical representations in the form of continuous magnitudes. Amongst them, we hypothesize that providing accurate information about the testing range, as well as feedback to the anchor values, are some of the critical aspects.

Some previous studies reported that providing adults with feedback about their estimates modifies the estimates throughout the tested range, even when the amount of feedback is minute (Feedback condition in Minturn & Reese, 1951; Krueger, 1984;

Izard & Dehaene, 2008). Likewise, informing participants about the response-range boundaries can reduce the amount of variability and increase accuracy (Krueger, 1984; Lipton & Spelle, 2005). Yet another evidence about the impact of reference trials on the calibration of the verbal estimates of adults comes from the “Numerosity Naming tasks”, in Revkin, Piazza, Izard, Cohen & Dehaebe (2008)’s study. After being informed about the range’s starting and endpoint numerosities, adults were trained intensively with accurate feedback, to quickly provide decade numbers estimates following the brief presentation (150-ms) of sets composed of 10 to 80 dots. Before each of the 40-trials testing blocks, participants were trained in 16 of these calibration trials. Mean estimates were accurate and, like in our results, the CVs increased and peaked at sample “30”, after which they decreased monotonically (Fig. 3.h from Revkin et al., 2008).

To conclude, we note that the presentation of nonverbal sets of dots may in fact tap into adult’s “number sense”. However, the explicit enumeration of the range boundaries, together with adults’ predisposition to offer multiples of ten as estimates, may have enhanced the (in this case, accurate) categorization of the sets. In other words, if one assumes that the adults arrive at the task already with the response bias to offer the verbal tags “ten, twenty, thirty, …, ninety” (e.g., Baird, Lewis, & Romer, 1970; Baird & Noma, 1975), the task may be solved on the basis of ordering the perceived numerical quantities (refer to Izard & Dehaene, 2008, and Dehaene, 2009, for a model that characterizes the effect of calibration, by separating effects originating from numerosity encoding and from response selection. The model assumes that subjects map segments of their internal log-Gaussian magnitude representation onto the requested response labels). An empirical test is still lacking, that may clarify under which procedural features is the use of one strategy favored over the other possible ones (e.g., magnitude vs. ordination vs. counting) (refer to Gandini, Lemaire, & Dufau, 2008, for more on response strategies when young adults are asked to enumerate large sets of dots, range [4,79]).

<u>Children</u>. The upper right histogram in Figure 11 shows that children offered the number words between <u>1 and 9</u> at approximately the same frequency (individual H scores averaged .98, SD = .04). Also, as their group curve suggests, mean estimates were fairly accurate but the number of errors tended to increase with the sample value. Likewise, standard deviation increased with the sample numerosities. Children did

make more errors than adults, whose performance was practically seamless. Overall, children's incorrect estimates represented 12% of the total responses. 72% of these errors were estimates that differed by ± one unit from the target numerosity. Additionally, errors of underestimation (i.e., the estimate was smaller than the target) were twice more likely than errors of overestimation (72% vs. 28%, respectively).

Except for the sample "1", where the standard deviation was zero, the group's coefficients of variation were constant across the numerosities (mean CVs = 0.12; with the linear regression yielding a slope of -0.0005, and an intercept at 0.1222). To our knowledge, only a few studies have addressed scalar variability in preschool children's estimation of small numbers, with the existing results being inconclusive. For instance, in a procedure where sample duration was 250 ms, Huntley-Fenner (2001) reported constant coefficients of variation across the numerical sets of {5,7,9,11} black squares. However, in a procedure where cards containing [1,10] black dots were presented during 1-s, Le Corre and Carey (2007; "Fast Cards" task) reported that the children's' estimates within the numerical range [1,4] did not show scalar variability. Whereas in both studies verbal counting was dismissed as a possible response strategy, only the former study attributed scalar variability to the mapping between number words and underlying approximate number representations. In view of their results, though, Le Corre and Carey (2007) excluded that children relied on analog magnitudes, and suggested instead that their estimations relied on another representational system with numerical content (but see Negen & Sarnecka, 2010, for a rebuttal and alternative explanation for Le Corre & Carey's results).

However, in our study no child followed the instructions to avoid counting the dots. Rather, all were witnessed counting the numerical sample, especially when it was a set of 4 or more dots. Often, children would touch the individual dots successively with their finger, while saying the sequence of number words (*tagging*, a term by Gelman and Gallistel, 1978, pp. 77-78). Therefore, in our case, scalar variability of the verbal estimates is purely a property of verbal counting responses and cannot be interpreted, as it usually is, i.e., as reflecting the mapping between number words and the variability of the underlying analog magnitude representations of numbers.

It is also difficult to compare our study with the remaining dot enumeration studies in similar numerical ranges. Not only are there just few of them, but also their main interest resides in the relationship between the children's reaction times – a variable we did not record – and the sample numerical value (e.g., sample range [1,8] in

Reeve et al., 2012; sample range [1,5] in Gray & Reeve, 2014). Namely, their purpose was to pinpoint a "subitizing range", that is, by observing the reaction times patterns, to identify up until which numerosity children can accurately enumerate without explicit internal or external counting (e.g., Starkey & Cooper, 1995; Simon & Vaishnavi, 1996; Benoit, Lehalle, & Jouen, 2004).

Inspection of the scatterplot Figure 11, regarding children's verbal estimation of sets of dots within the [10,90] range, shows that although group mean estimation increased monotonically with sample numerosity, there was great variability in these estimates, as indicated by the large standard deviation bars across the numerosities. Standard deviation increased from about 9.8 in sample "10" to 24.8 in sample "30", at which point stabilized and later on slightly decreased after sample "70" (resembling an inverted U shape). The CVs decreased linearly with sample numerosity (best fitting linear regression yielded R^2 = .99).

Again, our data lead us to defend our argument that the dominant source for such variability is not the numerical subjective scale, but strategies related to the number words known by the participant and/or the occurrence of counting behavior, including miscounts - skips and doublecounts (Gelman & Gallistel, 1978). For instance, the lower right H histogram from Figure 11 shows that the distribution of the estimates' frequencies was very uneven. Estimates from "one" to "thirty" were offered more frequently than the corresponding numerosities were actually presented, as indicated by the H bars' relative frequency in Figure 11 being over 0.11, at about 0.15. Number words between [31,80] were offered less frequently than the actual samples on this range (relative frequency of about 0.06 each bin). Finally, number words between [81, 90] were offered considerably more than any other equally sized bin (rel. frequency 0.28). More important, the same or similar patterns were also observed at the level of the individual histograms of estimates.

Figure 12 shows the distribution of the estimates' relative frequencies in more detail. The figure shows the just mentioned conglomeration of verbal estimates until around thirty. Moreover, these estimates were offered following all the numerical samples, even the largest numerical sets (as seen by green and blue dot series within the [1,30] domain range in the x-axis of Figure 11). This happened because children frequently ignored the instructions not to count and pointed at each dot, one after the other, while counting them aloud, until the set disappeared (after 20 s) or until they had individuated all the items. In this last case, the offered number depended on the limit of the child's number word series, which sometimes would be re-started once the child reached her maximum counting word with still to-be-tagged dots remaining. Whenever the trial's maximum time elapsed, the same set was repeated on the next trial; children would resume counting, up until they reached their number word series' limit and then offer this last word as a response. In another type of strategy, when the numerical sample was a larger number, children either "gave up" counting after *tagging* a few items or immediately offered an estimate. This estimate was either the largest anchor, "ninety" (about which they had been explicitly instructed), or another decade number. In fact, out of all the estimates within the range [31,90], more than 96% were decade number words, a much larger percentage than the observed for the estimates in the [1,30] interval (56.6% of which were decade words). Figure 12 illustrates how frequent were verbal "ninety" responses, following a wide range of numerical samples, particularly the three larger sample values. In other words, when they seemingly put off counting and offered an estimate, many children would offer the largest numerosity they had been told about, "ninety". Recall that the range was explicitly told before the task and throughout the test, feedback was provided for the anchor sets as well as number words. Therefore, it is possible that such prevalence of the largest anchor estimate is an "end effect" (Simon & Vaishnavi, 1996), in the sense that because children know the maximum number of dots that can appear, whenever they are unsure or they cannot enumerate a large set they offer this value.

Finally, it is noteworthy how the inspection of the group mean estimates (scatterplot in Figure 11) is a misleading summary of the preschoolers' performance on the Dots[10,90] verbal estimation condition. Curiously, mean estimates as a function of numerical sample was the main analysis carried out in Lipton and Spelke (2005)'s study, from whom we adapted our estimation task. Their study contrasted age and counting proficiency groups by fitting models and appraising the individual or group R^2 values (sample range [20,120] in Lipton & Spelke, 2005). Given how poorly individual and mean verbal estimation scatterplots describe our children's trial-by-trial performance, such analyses were not emulated in the present study.

3.4 Relation between number-to-position estimates, counting and verbal estimates

A final interest in our study was to investigate the relationship between number-to-position performance and counting or estimation abilities in preschool children. To that end, each child participant was classified in age (months), amount of pre-training in brightness (CtrlGroup – 0, ExpGroup – 1), use of response bar during the Number-to-position tasks (entropy scores: H_{-9}, H_{-90}), an Abstract Counting score (largest number word offered), a Counting Objects score (sum of correct responses – [0,27]), and a measure of the verbal estimates offered during the Verbal Estimation tasks (entropy scores: H_{V-9}, H_{V-90}). Table 3 presents the Spearman's rank order correlation coefficients between these variables.

Table 3

Spearman correlations between children's performance in the number-to-position, the counting and the verbal estimation tasks

Variable	PRE-train	Age	H_{-9}	H_{-90}	CtAbs	CtObj	H_{V-9}
PRE-train	–						
Age	.01	–					
H_{-9}	.64**	.06	–				
H_{-90}	.48*	.29	.64**	–			
CtAbs	.20	.38	.10	.34	–		
CtObj	.00	.69**	-.07	.18	.64**	–	
H_{V-9}	-.13	.33	.01	.27	.69**	.66**	–
H_{V-90}	-.34	.49*	-.10	.03	.53*	.67**	.64**

Note. PRE-train: absence or presence of a pre-training in brightness {0,1}; H_{-9} and H_{-90}: entropy scores of the spatial responses during the number-to-position Dots[1,9] and Dots[10,90] tasks, respectively; CtAbs: score in the Abstract Counting task; CtObj: score in the Counting Objects task; H_{V-9} and H_{V-90}: entropy scores of the verbal estimates.
* $p < .05$ ** $p < .01$

Age was found to be correlated significantly with the proficiency in counting objects as well as with the performance on the Verbal estimation [10,90]. Note that one would not expect the same effect to be verified with the Verbal [1,9] task due to all children's relatively accurate performance (*ceiling effect*). This result, suggests that the older the child, the more he/she tends to provide a wider range of number words. Together with our observation that many children engaged in the Verbal estimation

[10,90] task by attempting to count the samples, perhaps the older children present more estimation strategies. This is in line with the proposal that best counters also have more and better strategies to estimate discrete quantities (Crites, 1991; Gandini, Lemaire, & Dufau, 2008).

Besides the previously discussed relationship between the presence of a pre-training in brightness and the distribution of spatial responses across the response bar during the Dots[1,9] and Dots[10,90] conditions (see, e.g., Figure 7), no other variable was found to be correlated with the performance in the number-to-position tasks. Most important, whereas the scores from the three counting assessment tasks (Abstract, Objects and Verbal estimation) significantly correlated with each other, none significantly correlated with performance on the number-to-position tasks.

As we have previously discussed, many children's individual performance in the larger number-to-position task, was poorly characterized by simple linear and/or logarithmic models. But, for the sake of comparison with all previous studies, we repeated the correlations and further characterized number-to-position performance by the variance accounted for by the best fitting linear model (R^2_{lin}). Besides it being the traditional measure of linearity in numerical-spatial mapping studies, our interest in investigating R^2_{lin} was also to compare with the resulting conclusions from the analysis of Entropy scores (use of the response bar). Table 4 summarizes the results.

Table 4

Spearman correlations between children's linearity in the number-to-position task and performance in the counting and the verbal estimation tasks

Variable	PRE-train	Age	R^2_{-9}	R^2_{-90}	CtAbs	CtObj	R^2_{V-9}
PRE-train	–						
Age	.01	–					
R^2_{-9}	.30	.21	–				
R^2_{-90}	.23	-.21	.01	–			
CtAbs	.20	.38	.27	.20	–		
CtObj	.00	.69**	.35	-.09	.64**	–	
R^2_{V-9}	.00	.47*	.57*	-.02	.59**	.51*	–
R^2_{V-90}	.10	.43	.48	-.09	.70**	.58*	.57*

Note. PRE-train: absence or presence of a pre-training in brightness {0,1}; R^2_{-9} and R^2_{-90}: R^2 from the best fitting linear function of number-to-position Dots [1,9] and [10,90] tasks, respectively; CtAbs: score in the Abstract Counting task; CtObj: score in Counting Objects task; R^2_{V-9} and R^2_{V-90}: R^2 values from the best fitting lineal model of verbal estimates.
* $p < .05$ ** $p < .01$

R^2_{lin} values from the two Verbal Estimation tasks ([1,9] and [10,90]) were correlated with each other. In other words, the more linear-like a child's mean verbal estimates in the Verbal Estimation Dots[1,9] condition, the more linear-like the mean estimates are in the [10,90] range. Similarly to the Entropy measure, linearity in each of the verbal estimates tasks was correlated with the performance in the Abstract and Objects Counting tasks. Our main interest, nonetheless, relies on the relationship between these tasks and number-to-position performance. For the Dots[1,9] range, linearity of the positional estimates (number-line) was also correlated with the linearity of the verbal estimates. However, no significant correlation was found between the most critical variables: positional and verbal estimates in the Dots[10,90] range. That is to say, linearity in the number-to-position Dots[10,90] condition could not be predicted by the performance in the verbal estimation, abstract or objects counting tasks.

In the introduction of this study we mentioned that the majority of studies correlate number-to-position performance with the scores from Math achievement tests, and only a lesser number has collected evidence on how it relates to familiarity with numbers. In face of the correlational analyses in the current study, we found no evidence for a relationship between children's familiarity with numbers, counting or estimation abilities and their mapping of non-symbolic numerosities onto space. This occurred when number-to-position performance was evaluated both by the linearity of the child's estimate function (R^2_{lin}) and when we actually addressed the distributions of responses on the extent of the response bar (H scores).

We did find that linearity in number-to-position performance on the two numerical ranges correlate with each other, a finding that had also been reported in Berteletti and colleagues (2010)' study with kindergarten children, which were tested in symbolic 1-to-10 and 0-to-100 number-line tasks (Exp. 1). Most important, although in our study R^2_{lin} in the smaller number-to-position range was associated with R^2_{lin} in the larger range, linearity and Entropy scores were not related to performance in the other estimation and counting tasks. Therefore, our results contradict those from the first study that examined the effect of familiarity with numbers (Ebersbach et al., 2008). However, Ebersbach and colleagues' number-to-position procedure differed from ours in three main aspects. Firstly, children responded in a wooden external line, not in a bar presented on a screen. Secondly, the experimenter remained next to the child and at the beginning of each trial he was responsible for presenting the numerical sample. Thirdly, this sample had both symbolic (printed digits, spoken words) and nonsymbolic features (sets of dots), and not merely the latter as in our study. Which, if any, of these differences may explain why Ebersbach and colleagues found a correlation between familiarity with numbers and linearity in the number-to-position task, and we did not, we cannot answer without further exploration. It is possible that the accuracy of correlation coefficients was affected by both studies' small sample size (Schönbrodt, & Perugini, 2013), a fact that should caution us against over-interpreting the data.

Interestingly, though, the absence of a relationship between nonsymbolic number discrimination and counting abilities has also been reported in experimental procedures other than the number-to-position task (e.g., nonsymbolic number comparison, nonsymbolic numerical Stroop) (Soltész, Szucs, & Szucs, 2010; Rousselle, Palmers, & Noël, 2004; Rousselle, & Noël, 2008; Holloway & Ansari, 2009). Moreover, even the relationship between Math achievement scores and number-to-position or other estimation tasks does not seem as straightforward as it was initially speculated to be. For instance, recent data suggests that the relation between tasks is dependent on the young child being presented with either symbolic or nonsymbolic numerical stimuli (Sasanguie et al., 2013).

Since it seems unlikely that preschool children would respond solely on the basis of their psychological representation of numbers, future studies ought to probe on the possible influence of mediating behaviors such as self-made rules. Lastly, a more systematic analysis is needed to isolate the contribution of each of the many numerical abilities which are encompassed by the term "familiarity with numbers".

4. Conclusions

The current study aimed to investigate preschool children and adults' performance in nonsymbolic number-to-position tasks. Namely, participants were trained to respond at the endpoints of a response bar following numerical sets of 1 and 9 or 10 and 90 dots. Next, they were tested with intermediate numerosities. A review of the current literature, together with the results of our first study, had led us to expect that preschoolers do not respond along the response bar, at least not to the same extent as adult participants. In our former study we found that, instead of responding progressively towards the right side of the bar as a function of the sample numerosity, preschoolers distributed their responses in differentiated proportions at two or three locations of the bar ("categorical responding"). The problem with such categorical responding is that it violates the fundamental premise of the number-to-position task (i.e., mapping) which, in turn, hinders the contribution of these tasks as evidence towards the hypothesis that numbers are inherently mapped onto space (Núñez, 2011; Núñez, Cooperrider, & Wassmann, 2012). Moreover, our data illustrated that an average location should not be considered as characterizing the actual locations along the extent of the response bar and that the resulting log-like feature of average location was an analysis artifact.

In the previous mapping experiment, prior to completing the number-to-position task, all participants had been tested in a Brightness-to-position procedure. This pre-testing was implemented to ensure that children had been exposed to a task requiring them to respond along a spatial *continuum*. Nonetheless, many preschoolers still responded categorically when later tested in the number-to-position task. Because all participants experienced the pre-training in Brightness, we could not disentangle this possible contribution to the use of the response bar when participants were tested with numerosities. For this reason, in the current study we tested the effect of pre-training in a Brightness-to-position task, by separating participants into Control and Experimental groups. The Experimental group received a pre-training in the Brightness-to-position task, but this pre-training was different from that in our previous study, in that (i) it was adapted to approach as much as possible the numerical task, and (ii) there was explicit training (with feedback) of selecting positions further to the right of the response bar as a function of progressively darker stimuli.

The results of the present study showed that a continuous repertoire was established after training the mapping of nine separate values along the brightness dimension onto nine locations along the response bar. This Brightness-to-position pre-training enhanced responding along the bar in a Number-to-position task. When comparing the Control and Experimental groups, the percentage of preschoolers who distributed their responses into two or three clusters of locations decreased from 8.3% to 0% in the Dots[1,9] condition and from 37.5% to 8.3% in the Dots[10,90] condition. Likewise, an equivalent effect of the experimental condition was shown by the increase of the normalized entropy scores.

In the previous experiment, we had questioned whether tri-categorical mapping patterns could have been a by-product of the non-differential reinforcement following the intermediate numerosities. In the current study, however, no reinforcement schedule was in effect, there were no instructions about the midpoint mapping and the testing proceeded without feedback. And yet, tri-categorical patterns are also present. Perhaps, as Barth and Paladino (2011) propose, participants approach the number-to-position tasks as proportion judgments between two (the anchors sites), three (anchors plus the midpoint) or progressively more reference points. A noteworthy fact is that the tendency to concentrate responses at the bar's midpoint also appears in verbal number-to-position tasks (Barth & Paladino, 2011; Barth, Slusser, Cohen, & Paladino, 2011; Hollands & Dyre, 2000; Cohen & Blanc-Goldhammer, 2011; Sullivan, Juhasz, Slattery, & Barth, 2011; Ashcraft & Moore, 2012; though see Opfer, Siegler, & Young, 2011 for a rebuttal of the proportion-model). However, though they are the most frequently employed variations of the number-to-position tasks, Arabic digit studies have not yet tested how explicit instructions and/or the prior training about the middle numerosity – midpoint location mapping affect performance.

The question remains as to why continuous responding (indicated by H and k clusters scores) occurs more frequently in the smaller range (Dots[1,9]) than in the larger range (Dots[10,90]), as indicated both by H scores (Figure 7) and number of response clusters (Figure 9). According to the experimenter's notes, children would occasionally count the number of elements once the numerical sample was presented. These attempts to count the sample occurred more frequently during the Dots[1,9] condition. Assuming that a child would count the sample, even if not on all trials, it is possible that a few counting instances are enough to attribute numerical identity to each of the sample values. To provide a *tag* – in this case, a number word – to each set may facilitate spatial differentiation in the response bar. For instance, if the child estimates or counts a set as "5" and another set as "6", this *tagging* will facilitate the understanding that different values are being presented and the understanding of the instruction to place the numbers along the bar, in separate locations. That is, because "5" is different than "6", they cannot be placed on the same location. In the Dots[10,90] number-to-position task, after being trained in the anchors, the child started a testing with also intermediate values but it would be less likely for her/him to attribute a number word (or any other unique verbal tag) to each numerical sample. Enumeration or tagging would be made harder because of number discriminability. In other words, contrary to Weber's Law of number discrimination, because children can count 5 and 6 dots, but not 50 and 60 during the sample presentation, it may be harder to discriminate 50 vs. 60 than 5 vs. 6 (Trick & Pylyshyn, 1994).

Another setback for tagging in the Dots[10,90] condition is that most children do not know or offer the number words for the numerosities within this range, as our Verbal Estimation tasks revealed. As they observed the numerical arrays during the number-to-position procedure, some children would spontaneously provide verbal expressions such as "*very few*", "*some*", "*more or less*", "*a lot*", "*many*", or "*the most of all*", among others. These expressions are common verbal quantifiers, but their meanings are imprecise (Borges & Sawyers, 1974; Routh, 1994). It is also conceivable that children's performance in number-to-line tasks was mediated by the mappings between verbal quantifiers and the unidimensional scale of quantity (number of dots) (Crites, 1992; Routh, 1994; Moxey & Sanford, 2000; Siegler & Robinson, 1982; Laski & Siegler, 2007). Moreover, there is a reason to assume that verbalizations play a part in preschoolers' number discrimination, as shown in the inspection of individual performances in numerical bisection procedures (Almeida, Arantes, & Machado, 2007). A bisection procedure is a 2AFC task, where subjects have to respond categorically: following each numerical sample, select one of two response *manipulanda*: one for "few" and the other for "many" responses. However, the psychometric curve is supposed to be sensitive to numerosity, which is reflected in a *gradually* increasing proportion of "many" responses as a function of numerical sample. As a novelty in the literature on children's numerical discrimination, Almeida and colleagues (2007) examined the individual psychometric curves of preschool children, tested with sequences of 2 to 8 or 4 to 16 tones. Their most interesting finding was that some preschoolers always chose "few"-*manipulandum* after the first two smaller numerosities and always chose the "many"-*manipulandum* after the remaining numerosities, which resulted in abrupt Step-like curves. Also, the average group curve could not hint at the categorical pattern of responding observed at the individual unit of analysis. Such categorical responding appeared to be connected to the child's spontaneous verbalizations during the experiment, who would mention the quantity of tones and their self-made rules for the task. In conclusion, these results suggest that children verbally categorize the stimuli, so the "elementary" sensitivity to number may be replaced under some circumstances by behavior that is verbally mediated and that is expressed in the Step-like functions. Perhaps similar verbal classification and categorization strategies were the basis for the bi- and tri-categorical responding in our number-to-line procedures. If reliance on the verbal quantifiers induces categorical patterns of responding, then we could expect them to occur more frequently on the larger Dots[10,90] condition than in the Dots[1,9] one, because for the latter there are actual number words available as precise verbal quantifiers. Again, we would need more than the experimenter's notes, and a video-recording of children's sessions would have helped us to identify not only their patterns of responses, but also help inspect their verbalizations during the trials, and to pinpoint all the occasions when they counted the sample set.

Notwithstanding the possible effect of verbalizations, the Brightness-to-position pre-training still enhanced continuous responding along the bar in both ranges. Probably, young children's responding is the product of many strategies and influences, both related to the child's understanding of the instructions, experience with spatial continuous repertoires, or familiarity with numbers, among many other factors.

The current study provided yet another evidence that preschoolers' performance, in contrast with adults', is highly variable and tends towards categorical patterns. For many children, both the group and individual average location curves were misleading as to the trial-by-trial response locations selected by the participant. Our data was analyzed at increasingly finer units of analysis: group average, individual average and simple curve fitting and, lastly, single trials. The first two, which incidentally constitute the only levels appraised in former studies, were not able to detect and describe individual differences. Notably, inspecting solely the mean group curves would suggest a magnitude effect: linear mappings in the smaller [1,9] range and a logarithmic mapping in the larger [10,90] range. In addition, only the S-shaped median curve from the Control group hinted at the occurrence of categorical responding. Moreover, the within-participant comparison of R^2 values, taken as a measure of goodness of fit, found significantly good fitting of linear and logarithmic models. As such, these two units of analysis were unable to detect and filter non-continuous cases. Such endeavor was possible only at the trial-by-trial analysis. In conclusion, a key-point of our work was the case of how visualizations, i.e., more exploratory graphical methods (histograms and scatterplots), are needed to discover structure and patterns in the data (cf. Anscombe, 1973).

The evidence points for there being no unique associations between number and space perception that may be revealed in number-to-position tasks. It does not seem to be a matter of inherent mappings. Rather, we believe the number-line intuition is learned or, in other words, it does not preexist the learning of other cultural tools (Núñez, 2011; Núñez, Cooperrider, & Wassman, 2012). At the end of the previous study the following question had been posed: “if categorical responding is the “a priori” response pattern, then (…) under which conditions (prior learning experiences) does responding in a continuous one-dimensional space first become under the control of the (continuous) numerical stimulus dimension?” The current study moved one step in the direction of clarifying such conditions by showing that training for a spatial continuous response to be under control of a non-numerical continuous dimension might be one way to enhance the mapping of numbers onto space. The focus should continue to be on the acquisition of a numerical-spatial mapping, which should be treated as more akin to a motor-skill learning than to a preexisting intuition. By assuming this approach, one can venture that the similarities between the two tasks – brightness and number - may have led to positive transfer of learning (Mazur, 2002, p. 312; Adams, 1987; Schmidt, 1991). Our next step forward, thus, would be to determine which features of the current Brightness-to-position pre-training are relevant for transfer.

CHAPTER IV : STUDY 3

Control of a response continuum by the numerical stimulus continuum: isolating the effect of training responses in different locations of the bar

1. Introduction

The previous two studies have tested preschool children and adults in number-to-position tasks, with nonverbal numerosities as the samples. Crucially, the response location curves produced by averaging the single-trial data did not reflect preschoolers' performance (see Stoddard & McIlvane, 1989, for a similar discussion within the field of intradimensional discrimination training in young children). On the one hand, and similarly to previous authors' number-to-position procedures, the smooth average group curves increased monotonically, which suggested that as numerosity increases, responses occur further to the right of the bar. Moreover, the comparison between preschoolers and adults' group curves revealed a logarithmic-to-linear developmental transition, a finding that agreed with the current main model of number representation, the "Mental Number Line" (Dehaene, Bossini, & Giraux, 1993; Zorzi, Priftis, & Umiltà, 2002; Hubbard, Piazza, Pinel, & Dehaene, 2005; Opfer & Siegler, 2007; Dehaene et al., 2008, 2009; de Hevia & Spelke, 2009, 2010). On the other hand, inspection of single-trial scatterplots showed that many preschoolers responded categorically. By "categorically" we mean that these children restricted their responses to two or three sites of the response bar. These response clusters, that is, the locations on the response bar where responses were concentrated, mostly corresponded to the bar's endpoints and midpoint. And even an individual averaging artifact may occur at the individual level of analysis, when probability of selecting the rightmost cluster increased as a function of sample numerosity.

Given the data we have collected thus far, we remain skeptical of the existence of an innate association between number and space (i.e., the mental number line), as it may be revealed in standard number-to-position procedures carried with preschool participants. If one is to abandon the hypothesis of spontaneous number-space mappings, then we need to identify which experiences lead to this behavior. In view of our results, we propose that the differences between the assignment of smaller and larger numerosities into a spatial medium may be related to the degree that *tagging* occurring concomitantly with sample presentation. The most customary case of such *tagging* behavior would be counting the numerical set. In addition, we believe that when the numerical stimuli cannot be tagged easily with a number word, as in the case of larger numerosities, then familiarity with a continuous topography of response may be the determinant aspect for numbers to be mapped continuously.

For this reason, we will continue our study of preschoolers' mapping of numbers onto space by characterizing and manipulating the pre-training history. In the second study of the current thesis we found that, if prior to the number-to-position test preschoolers are trained to respond in different locations as a function of a sample's brightness, the tendency to respond along the bar during the numerical task improved. The similarities between the Brightness pre-training and the Number testing situations must have been the basis for the positive transfer (Thorndike & Woodworth, 1901; Perkins & Salomon, 1992; Lydersen & Perkins, 1974; Brown & Kane, 1988).

The reinforcement and feedback contingencies in our Brightness pre-training protocol probably have shaped behavior in the direction of the targeted pattern of responding, that is, for responses to be placed along the extent of the response bar. All participants learned the Brightness-Position assignments. This is the same as to say that they learned a continuous repertoire, given that "for each point along a stimulus dimension, a unique corresponding response was made along some response dimension" (Scheuerman, Wildemann, & Holland, 1978; Holland & Skinner, 1961; Rosenberg, 1963).

Which key features led to transfer to the numerical testing? The brightness pre-training shared two features with what children were later instructed to do in the numerical task (i.e., map increasingly larger stimuli values onto increasingly rightmost positions along the response bar). First, there are the features pertaining to the sample, in that the samples belong to a sensory *continuum* or, at the very least, they may be ordered according to their magnitude (for more information about generalization and intradimensional discrimination training in young children see, e.g., Landau, 1968, 1969; Stoddard, McIlvane & de Rose, 1987). Second, there are features pertaining to the response topography: the manual operation of selecting different sites on the response bar.

In the current study we have tried to isolate the effect of this second feature on a Number-to-position task. It could be that those children in the Control group, who were not taught to respond along the bar as a function of brightness stimuli, did not respond along the bar in the Numerical task because they had never responded in ordered positions. The verbal instruction to do so when tested with numerosities could have been made clearer to them had they ever touched the response bar at other places than the anchor sites.

Perhaps a simpler pre-training where the manual component - to touch the response bar (spatial *medium*) - is singled out, will be enough for the child to understand that, when presented with the number-to-position task, more than just the anchor positions can be selected along the bar. Hence, in the current experiment children in the Experimental Group will experience a pre-training demanding them to select different locations along the response bar as a function of a set of cartoon images. Critically, although during this Figures-to-position pre-training the stimuli are all cartoon TV figures they are not ordered along a *continuum*, that is, there is no magnitude or ordinal scale.

Finally, we will once again address the possible influence of familiarity with numbers and number-to-position performance (Lipton & Spelke, 2005; Le Corre & Carey, 2007; Ebersbach, Luwel, Frick, Onghena, & Verschaffel, 2008). To that end, the last experimental sessions will consist of screening preschoolers' counting and verbal estimation of arrays of dots.

2. Method

2.1 Participants

The experiment studied twenty-four Portuguese pre-schoolers (12 girls) and twenty-four Portuguese young adults (12 women). The mean age of preschool children was 4.86 years (SD = 0.60; range 4.17 - 5.85) and the mean age of adults was 21.62 years (SD = 2.53; range 18.38 - 31.37). Adult participants were Psychology undergraduate students from the University of Minho (Portugal), who volunteered to participate in the experiment in return for course credit. Informed consent was given by all adult participants and by the children's parents.

2.2 Numerosity stimuli

The same numerical stimuli that were created for the Dots[10,90] condition in Study 2 (see Methods section) were used in the current study.

2.3 Procedure

Participants were seated in front of a touchscreen laptop, in a separate room of the school. This was the same computer used during Study 2 (HP Pavilion tx2000 Notebook PC, screen size 12.1", screen resolution 1024 x 768, refresh rate 60 Hz). The previous (Study 2) experimental program in Visual Basic language was used to control all session events and record participants' responses in numerical tasks. A new experimental program, also in Visual Basic, was written for the sessions of the Pre-training condition.

The experimental conditions and the number of experimental sessions depended on the experimental group the participant was assigned to (Figure 1). The assignment of participants was carried as described in Study 2, ensuring that the groups were matched for sex and age. Participants in the Control Group were solely tested in a number-to-position mapping task. Those in the Experimental Group were trained in a line mapping task with images of five cartoon characters, and then were tested in the numerical mapping task,. At the end of the experiment, all participants were assessed in counting and verbal estimation tasks.

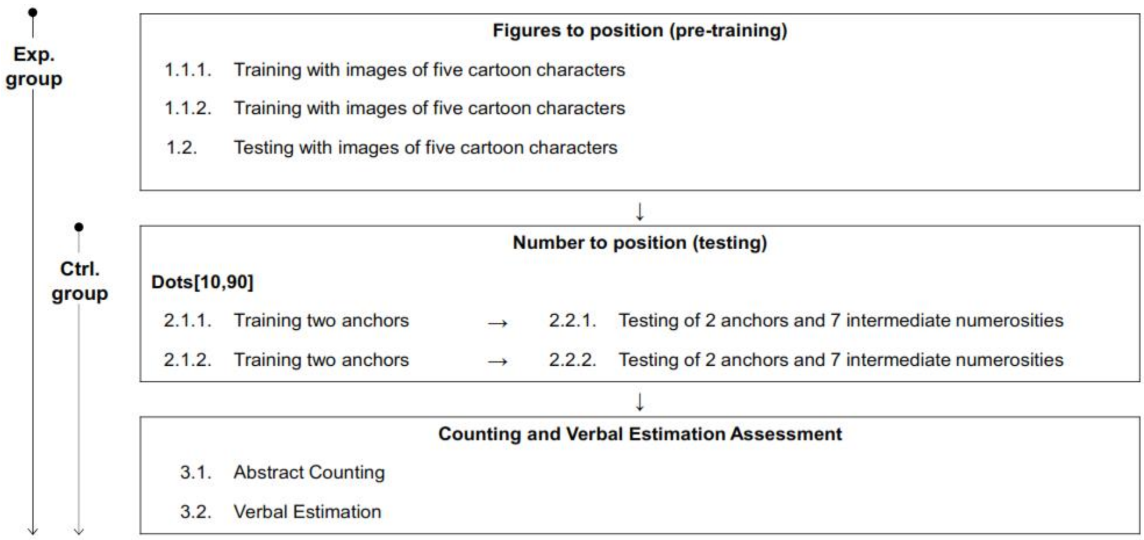


Figure 1. Diagram depicting the progress through the three experimental phases: Figures to position (pre-training), Number to position (testing) and Counting and Verbal assessment. Participants in the Experimental group underwent the Images to Position pre-training before being tested in the Number to Position tasks, whereas Control Group participants started from the Number to position tasks onwards.

1. Figures to Position (Pre-training). During the first experimental phase, participants in the Experimental Group learned to respond in a spatial continuum as a function of the images of cartoon characters. This Figures-to-position pre-training aimed to teach participants to orderly map visual stimuli onto space when, contrary to the numerical task, the samples are not part of a stimulus dimension *continuum* and may not be ordered by increasing magnitude. Therefore, this pre-training meant to isolate the "mechanical" effect of having differentially responded along a continuous space dimension.

1.1. Training five figures. At the beginning of the session, participants were presented the response bar, which was a uniformly colored yellow rectangle, approximately 26 cm width x 1.4 cm height, located 12.5 cm below the upmost part of the screen. In addition to the response bar's configuration, most procedural features were kept from Study 2: (i) the trial started with the screen colored in cornflower blue; (ii) the ITI lasted 1.5-s; (iii) after the ITI a star image appeared and a touch to it ("start response") triggered the appearance of the sample stimulus.; (iv) the sample stimulus was displayed horizontally centered and located about 0.1 cm below the upmost part of the screen; (v) both the sample stimulus and the response bar remained on screen until a touch at the bar of after 20-s without a response; (vi) the location of the first touch within the response bar was signaled by a thin dark blue vertical strip appearing and bisecting the response bar; (vii) the program saved choice latencies and the absolute and relative x/y coordinates within the response bar; (viii) a correct response was followed by verbal feedback; (ix) an incorrect response triggered the correction procedure, which was a two-step loop of a first experimenter-guided trial followed by a second, experimenter-independent trial.

The differences from Study 2's pre-training protocol pertained to the nature of the sample stimuli. Prior to the computerized task, participants observed the yellow bar depicted on the screen and handled printed cards with the images of five cartoon characters, which are depicted in Figure 2. The five cartoon characters were selected because children participants used to watch these cartoon series during the school break hours. Thus, they entered the experiment already possessing the correct mappings between images and verbal tags. This knowledge was confirmed when children were handling the printed cards.

'Shaun the Sheep'	'Donald Duck'	'Noddy'	'Winnie-the-Pooh'	'Tinker Bell'

Figure 2. Images of cartoon characters used as sample stimuli in the Figures-to-Position trials. Though for illustration purposes we here selected just one image of each cartoon character, a novel image exemplar was presented at any new trial.

Participants received the instruction that the yellow response bar was a path and that the cartoon characters had to walk this path starting from the beginning (while saying this, the experimenter pointed to the bar's leftmost corner) and further to the right until the path's end (pointing at the bar's rightmost corner). The experimenter said that some characters had to walk more than others. For instance, 'Shaun the Sheep' stayed at the beginning of the path (pointing to the leftmost position), 'Donald Duck' went a little further, 'Noddy' even more, 'Winnie-the-Pooh' even further than 'Noddy', and 'Tinker Bell' walked the path until its endpoint. While providing this verbal instruction, the experimenter moved her finger along the length of the response bar, in the left-to-right orientation, and touched it at five equidistant locations.

Next, participants started the computerized training procedure. During this phase, both verbal feedback and the correction procedure were in effect. Participants underwent two sessions of training with the five cartoon characters as sample stimuli. Each session comprised eight novel images of each cartoon character, randomly presented, resulting in a total of 40 trials per session. There was at least half a day of interval between consecutive experimental sessions. At the end of each session, the child participant received a sheet of stickers.

1.2. Testing five figures. After training, participants underwent one testing session. 14 novel images of each cartoon character were presented, which amounted to 70 trials in total. The order of figure presentation was randomized across trials. During the testing phase, the experimenter did not provide feedback, nor was the correction procedure in effect. After each response to the bar, the vertical blue strip signaled the selected location and remained on screen during 1000 ms, after which the ITI started. At the end of the session, the child participant received a sheet of stickers. An interval of half a day separated the next experimental phase, the Number-to-position task.

2. Numbers to Position (testing). Participants from the Experimental group progressed to the testing phase with Dots[10,90] Number-to-position sessions. As for participants in the Control Group, this was their first experimental phase.

All participants were initially trained to select the response bar's endpoints following sets of ten and ninety dots, i.e., the mappings 10-leftmost and 90-rightmost. Next, they were tested with both these and seven intermediate numerosities. Each participant emitted 16 responses per each numerical sample. All procedural details remained as in Study 2.

3. Counting and Verbal Estimation assessment. Children were tested on their Abstract Counting proficiency. Lastly, both children and Adults participants underwent verbal estimation sessions, with dots in the [1,9] and [10,90] ranges. The procedures were as described in Study 2's Methods section.

3. Results and discussion

3.1 Figures Pre-training

All child and adult participants in the Experimental Groups learned to select five regions of the response bar, as a function of the images of the five cartoon characters. Figure 3 depicts the mean response locations along the bar. The cross data points show the performance during the two training sessions. We considered the first response emitted by the participant, regardless of whether a correction trial followed it, or not. As illustrated by the cross data points, which fall at the corresponding grid intersections, participants learned to respond along the bar.

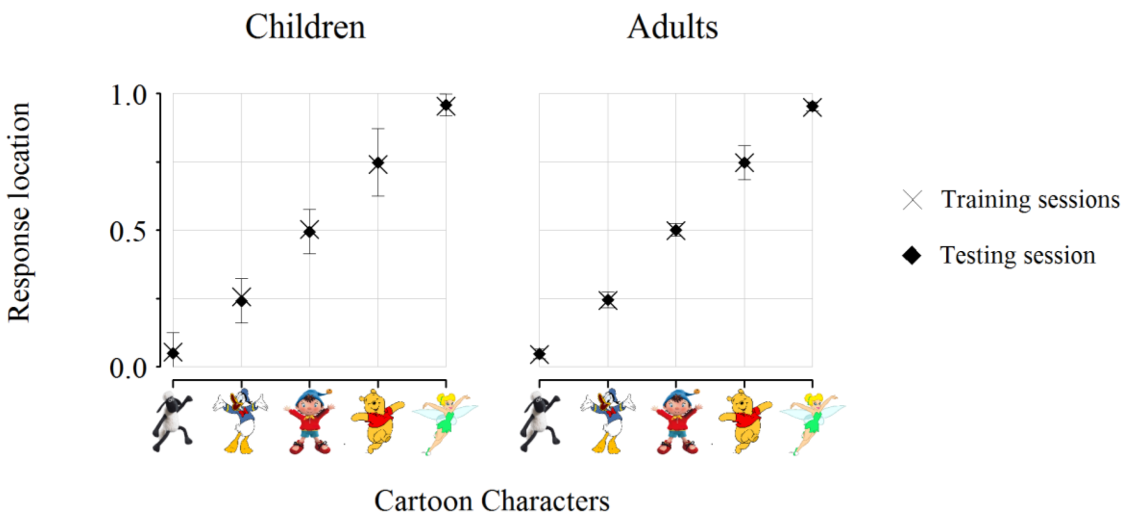


Figure 3. Response locations during the Figures-To-Position Training and Testing phases, for the children and adults' Experimental Groups. The cross data points represent mean response locations (± 1 SD) during the two training sessions and the filled diamonds represent the testing session.

In Figure 3, the results from the test session are depicted by the filled diamond data points. On test trials, without feedback and correction procedure, performance remained as accurate as during training. Appendix A contains each subject's scatterplot, which depicts single trials' responses during the testing session. Visual inspection of these scatterplots corroborates that all participants in the Experimental Group, children and adults, differentially responded along the bar as a function of the cartoon images.

Participants' successful performance on this pre-training has ensured that, especially in the case of children, they had experienced a task demanding a continuous response dimension. Our hypothesis was that having to master this pre-Figures-to-Position pre-training, especially in the case of children, would lead to a larger number of children responding in a continuous pattern when presented numerical samples, in comparison with those participants in the Control Group.

3.2 Numbers to Position (testing)

3.2.1 Group analyses

The two panels of Figure 4 depict the data from the Experimental and Control groups from adult participants; the right panels depict children's results. In addition, within each age group, the leftmost graph depicts the group's mean location of responses, and the right graph the median. For both children and adults participants, response locations increased as a function of the numerosity presented.

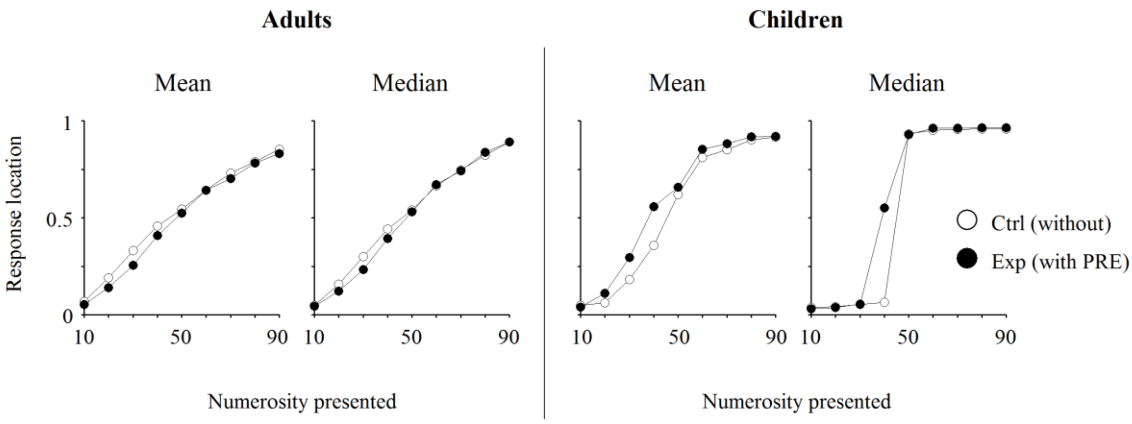


Figure 4. Numerical mapping of children and adults, separated by experimental group. For each age group, the y-axis of the graphs on the left column depicts the mean location of responses, from all trials taken together. The graph in the right column depicts the median of these trials. Within each graph, the filled dots depict responses from the Experimental Group, which had previously experienced the Figures pre-training, and the white dots the data from the Control Group.

Adults. Visual inspection of adults' group scatterplots, depicted in the left portion of Figure 4, shows that response locations increase as a function of numerical sample (significant Pearson correlation coefficients, $r(7) = .99$, $p < .001$, in both groups' curves, when one attends either to the mean or median response variable). The trained anchor mappings were not completely preserved during testing. Although the curve shows that the leftmost position was selected following the smallest numerosity, mean response location for the largest numerosity was 0.85 in the Control Group and 0.83 In the Experimental Group (medians were .89 on both). This resulted from, at the individual level, 7 participants on each group who on average responded at bar locations under .89 for the largest anchor.

The visual inspection of the curves does not allow for a decision between a more linear- or logarithmic-like spatial mapping. Also, the two group curves almost superimpose, be it the mean or median location graphs. Lastly, there are no striking displacements between the mean and median curves. Ordinary least squares (OLS) multiple regression analyses were carried out to contrast the contribution of a non-linear component (β_{log}) over and above a linear regressor (β_{lin}). These were the same analyses implemented in Dehaene et al. (2008) and Núñez et al. (2011)'s number-to-position studies with Western adults. Concerning the median and mean group curves of the Control group, both the linear and the logarithmic regressors are significant, but the logarithmic ones present larger weights (mean: $\beta_{log} = 0.350$; median: $\beta_{log} = 0.371$; see Appendix B for the full results of the multiple regression analyses). As for the mean and median group curves of the Experimental group, though the logarithmic regressors present larger weights as well (mean: $\beta_{log} = 0.246$, p = n.s.; median: $\beta_{log} = 0.168$, p = n.s), only the linear regressors are significant (mean: $\beta_{lin} = 0.008$, $p = .003$; median: $\beta_{lin} = 0.009$, $p = .001$).

The fitting of simple linear ($y = m.x + b$) and logarithmic ($y = m.\log_{10}(x) + b$) functions to the mean response location curves revealed marginally superior coefficients of determination (R^2) by the linear function. R^2_{lin} was.98, whereas R^2_{log} was .96 and .94 in the Control and Experimental groups, respectively. The same analysis applied to the median response location confirmed larger R^2_{lin} values ($R^2_{lin} = .99$ vs. $R^2_{log} = .95$ and .92 in the Control and Experimental groups, respectively).

We thus found that different type of analyses on the group curve led to different conclusions. On the one hand, although adults' group data was well described by either of the two models (linear and logarithmic), simple curve fittings suggest that mean and median response locations are better described by a linear function. On the other hand, results from the multiple regression analyses suggest that numerosities in the [10,90] range tend to be mapped logarithmically, rather than linearly.

Curiously, in both Dehaene et al. (2008)'s and Núñez et al. (2011)' studies which tested North American adults in similar stimulus conditions (simultaneous arrays of dots), the mappings were preferably logarithmic. However, that was not what we found in our Study 1, and neither did Anobile et al. (2012) in theirs. In these last two experiments, adults mapped larger arrays of dots in a more linear-like fashion. As we discussed in Study 1 and Study 2, the different results between number-to-line tasks with larger numerosities do not seem to be due to methodological features such as sample and response duration, or even the controls implemented for the non-numerical confounds co-varying with numerosity. Our provisional hypothesis remains that enumeration of the numerical sets occurring concomitantly with sample presentation may play a moderating role in number-to-position performance.

As for differences between experimental groups, the OLS multiple regression and the simple model fitting analyses suggest a weak effect of the experimental manipulation in the direction of a more linear-like pattern of spatial responding.

Children. In the right portion of Figure 4, presenting children's group data, one observes that both the mean and median curves of the Control and Experimental groups reach up to the largest anchor mapping (i.e., to positions further than .89). Similar to adults' curves, the mean location curves resemble a continuous pattern in the sense that, as numerosity increases, response locations increase monotonically. Contrary to adults' curves, however, the median curves behave distinctly from the mean ones. Children's median curves resemble a categorical, step-like function type of behavior. This is specially the case of the Control group's median curve, where responses occur practically only at the endpoints of the response bar. Accordingly, the OLS multiple regression analyses revealed that, though β_{log} regressors are larger than the β_{lin} ones, no regressor reached statistical significance (all p values are larger than .05) (Appendix B).

The results from the fitting of simple linear and logarithmic functions are not as straightforward (Tables 1 and 2). First, all four group curves were significantly fit by the two functions. Second, either when one attends to mean or median location, the Control group's curve is better fit by the linear function (mean: R^2_{lin} = .94 and R^2_{log} = .88; median: R^2_{lin} = .77 and R^2_{log} = .70). The differences between models regarding the

Experimental group's curves, however, are less striking than in the Control group. For instance, the R^2_{lin} and the R^2_{log} for the Experimental group's mean curve are .92 and .94, respectively. As for the Experimental group's median curve, both R^2 are .81, with the absolute difference between R^2_{lin} and R^2_{log} being less than seven thousandths.

Instead of following previous' authors method of assigning/describing the data in terms of number of logarithmic or linear cases according to the best fit, we would rather point out that children's group graphs in the current study highly suggest categorical patterns of responding at the individual level.

A last point concerns the effect of the experimental manipulation. When visually comparing the two group curves, the curve of the Experimental group is less steep than the curve of the Control group. This visual inspection suggests an effect of the pre-training with Figures-to-position mappings. In other words, the "mechanical" feature of the pre-training, i.e., responding along a continuous dimension as a function of arbitrary images, seems to have impacted children's performance when they were asked to position numbers in a same continuous response dimension. And yet, the question is to what is this effect translated into? Inspection of the group curves is not sufficient because even the Experimental group curves suggest categorical patterns of responding, especially in the median curve. As such, we cannot discard the possibility that even the gradual Experimental group curve is also an averaging artifact resulting from individual categorical responding. Moreover, it is possible that rather than enhancing responding alongside the bar, the pre-training increased the occurrence of the midpoint mapping responses (tri-categorical mapping), thus the midpoint mapping that occurs for sample '40' in the median's graph. Perhaps a better way to pose the question on the effect of the pre-training is to ask if in two equally sized and age and sex-matched samples, those who experienced pre-training present lesser number of cases of anchor-restricted responding (bi-categorical patterns).

Irrespective of what we will further discuss regarding individual patterns, attending to children's mean and median curves, again we found evidence that the averaging measure one attends to may drastically change the conclusions. This was an issue that arose during Study 2's discussion and, as we have then remarked, the group level of analysis does not inform us about the features of the individual patterns. In the current study, however, the median curves strongly suggest that many children have responded in a non-continuous fashion, more specifically a bi-categorical pattern.

3.2.2 Individual analysis - simple curve fitting

As was the case with the group level of analysis, simple linear and logarithmic functions were fit to each individual's mean and median response location curves (Tables 1 and 2).

Table 1

R^2 values yielded by the fitting of simple linear and logarithmic functions to each participant's *mean* *response location and to the group mean curve.*

	Children				Adults			
			Dots [10,90]				Dots [10,90]	
			R^2				R^2	
		Age	Lin	Log		Age	Lin	Log
Ctrl.	**g5**	4.17	.86	.65	**m19**	19.06	.96	.91
	b4	4.21	.94	.83	**m24**	19.12	.99	.89
	b11	4.43	.94	.86	**w23**	19.16	.96	.97
	g10	4.43	.83	.88	**w20**	19.28	.82	.98
	g7	4.48	.89	.75	**w15**	20.55	.97	.92
	b15	4.56	.88	.83	**m21**	20.77	.98	.91
	g12	4.75	.90	.86	**w16**	21.02	.95	.97
	b16	4.84	.84	.89	**w18**	22.02	.94	.99
	g20	5.47	.90	.90	**m17**	22.44	.98	.95
	g9	5.67	.88	.86	**w4**	23.12	.99	.93
	b24	5.73	.95	.80	**m14**	23.52	.98	.90
	b21	5.85	.86	.86	**m22**	31.37	.96	.98
	Avg	4.88	.89	.83	**Avg**	21.79	.96	.94
	SD	0.62	.04	.07	**SD**	3.41	.05	.04
	Group		.94	.88	**Group**		.98	.96
Exp.	**b2**	4.17	.86	.85	**m11**	18.38	.98	.97
	g3	4.20	.92	.97	**w5**	20.54	.95	.95
	b6	4.45	.73	.84	**w2**	20.60	1.0	.91
	b13	4.45	.91	.78	**w12**	20.74	.95	.96
	g1	4.46	.99	.91	**w7**	21.19	.92	.93
	g8	4.48	.84	.76	**m3**	21.47	.97	.91
	g22	4.62	.95	.90	**m10**	21.49	.99	.95
	b23	4.89	.80	.86	**w1**	21.99	.98	.94
	g14	5.26	.94	.93	**m13**	22.57	.98	.93
	g18	5.51	.90	.94	**m8**	22.65	.97	.92
	g19	5.82	.63	.86	**w9**	22.72	.98	.86
	b17	5.84	.86	.90	**m6**	23.13	.97	.93
	Avg	4.85	.86	.87	**Avg**	21.46	.97	.93
	SD	0.61	.10	.06	**SD**	1.31	.02	.03
	Group		.92	.94	**Group**		.98	.94

Table 2

R^2 values yielded by the fitting of simple linear and logarithmic functions to each participant's median response location and to the group median curve.

	Children				Adults			
			Dots [10,90]				Dots [10,90]	
			R^2				R^2	
		Age	Lin	Log		Age	Lin	Log
Ctrl.	**g5**	4.17	.63	.43	**m19**	19.06	.96	.88
	b4	4.21	.84	.72	**m24**	19.12	1.0	.89
	b11	4.43	.76	.60	**w23**	19.16	.94	.95
	g10	4.43	.68	.75	**w20**	19.28	.79	.97
	g7	4.48	.75	.59	**w15**	20.55	.95	.89
	b15	4.56	.77	.71	**m21**	20.77	.99	.91
	g12	4.75	.76	.70	**w16**	21.02	.94	.97
	b16	4.84	.68	.74	**w18**	22.02	.92	.98
	g20	5.47	.76	.70	**m17**	22.44	.97	.94
	g9	5.67	.75	.69	**w4**	23.12	.98	.93
	b24	5.73	.91	.74	**m14**	23.52	.99	.89
	b21	5.85	.79	.80	**m22**	31.37	.93	.96
	Avg	4.88	.76	.68	**Avg**	21.79	.96	.91
	SD	0.62	.07	.10	**SD**	3.41	.02	.04
	Group		.77	.70	**Group**		.99	.95
Exp.	**b2**	4.17	.68	.75	**m11**	18.38	.98	.96
	g3	4.20	.89	.91	**w5**	20.54	.96	.93
	b6	4.45	.68	.74	**w2**	20.60	.99	.85
	b13	4.45	.76	.60	**w12**	20.74	.95	.95
	g1	4.46	.95	.86	**w7**	21.19	.92	.92
	g8	4.48	.75	.59	**m3**	21.47	.96	.91
	g22	4.62	.89	.87	**m10**	21.49	.99	.94
	b23	4.89	.68	.75	**w1**	21.99	.99	.94
	g14	5.26	.90	.88	**m13**	22.57	.99	.91
	g18	5.51	.79	.86	**m8**	22.65	.95	.90
	g19	5.82	.53	.73	**w9**	22.72	.96	.81
	b17	5.84	.68	.74	**m6**	23.13	.95	.89
	Avg	4.85	.77	.77	**Avg**	21.46	.96	.93
	SD	0.61	.12	.11	**SD**	1.31	.06	.04
	Group		.81	.81	**Group**		.99	.92

Adults. All adult participants were significantly fit by both models ($p < .05$). Mean response location curves of 5 out of 12 participants from the Control group were better fitted by a logarithmic model (w16, w18, w20, m22, and w23), and the remaining 7 by the linear model. The number of "linear" participants in the Experimental group increased to 10, with only 2 participants being classified as "logarithmic" (w7 and w12) (refer to Table 1 for the complete individual scores). The log vs. lin comparison, now applied to median response location, confirmed the larger number of linear cases in the Experimental group (refer to Table 2 for the complete individual scores). The same participants from the Control group kept their classification. In the Experimental group, however, all participants were now classified as "linear" cases.

Thus, the number of participants categorized as "linear" in the Experimental group was always superior to those in the Control group (10 against 7 when attending to mean location; 12 against 7, attending to the median). Contrasting R^2_{lin} coefficients in the two groups, we found that the average of individual R^2_{lin} values was also higher in the Experimental group. However, these differences are not statistically significant (independent samples t-tests found no significant difference when attending to the mean: $t(1,22) = -.91$, $p = .37$; or median response location curves: $t(1,22) = -1.11$, $p = .28$). Because of that, we conclude that the pre-training did not lead to an increase of linearity in the placement of responses along the response bar.

Children. All children's curves were significantly fitted both by the linear and the logarithmic models ($p < .05$). Four children's mean mappings in the Control group were better described by the logarithmic model. These were participants g10, b16, g20, and b21 but in these last two cases, the difference between R^2_{lin} and R^2_{log} was merely on the order of the thousandths. In the Experimental group, participants were evenly distributed between the logarithmic and linear models (i.e., 6 cases on each model). As for the median data analysis, again the number of linear cases in the Control group was larger than logarithmic ones. Three participants from the Control group (g10, b16 and b21) presented larger R^2_{lin} than their R^2_{log} coefficients. As for the participants of the Experimental group, as it happened with the mean measure, there were half on each of the models. A mixed between-within ANOVA with experimental group (Control and Experimental) as the between factor and the R^2 yielded by the fitting of the model (Lin and Log) to the individual means as the within factor, revealed no significant effect of experimental group ($F(1,22) = .119$, $p = .733$), of the fitted model ($F(1,22) = 1.499$, $p =$

.241) and a marginally significant interaction between the two ($F(1,22) = 3.905$, $p = .061$). The same statistical test applied to the fitting of the individual median curves also only found an interaction effect ($F(1,22) = 4.452$, $p = .046$).

Children's results are quite unexpected because it was the children who experienced the pre-training with Figures-to-positions who presented more logarithmic-like numerical mappings, in comparison with the Control group. Did the pre-training disrupt the linearity of the number-space associations? As one might expect given the data we collected thus far, this is a deceptive question. The reasons are twofold. First, the individual fitting of functions and the subsequent dichotomous classification into "linear" or "logarithmic" cases are typically carried out, despite the possibility that some participants might present identical R^2_{lin} and R^2_{log} coefficients. This was the case, both with children and adult participants, in the current as well as in the previous study, which addressed R^2 values at the individual level (for all R^2_{lin} and R^2_{log} values, refer to Tables 1 and 2).

The second, more critical reason concerns the relationship between average curves and single trial responses. In Figure 5, we present single trial and average data from three children which illustrate this second point. Again we found that significant and relatively high R^2 values resulted from the fitting of children participants when the average curve behaves distinctly different from the fitted model. To illustrate our point, notice, for instance, that participants b11 and g14 present similar R^2 values regarding their mean location of responses. However, inspection of single-trial responses reveals that b11 restricted his responses to the bar's endpoints, with an increasing propensity to select the rightmost endpoint as a function of sample numerosity, which in turn resulted in increasing average curves. In contrast, inspection of participant g14's single-trial scatterplot shows that locations other than the endpoints were selected. Note that this pattern is not yet an exemplary "continuous" pattern, but one can observe that there are about four clusters of responses, with g14 touching the bar at the left endpoint, slightly further to the right from this anchor site, around the midpoint and, lastly, at the rightmost endpoint (largest anchor).

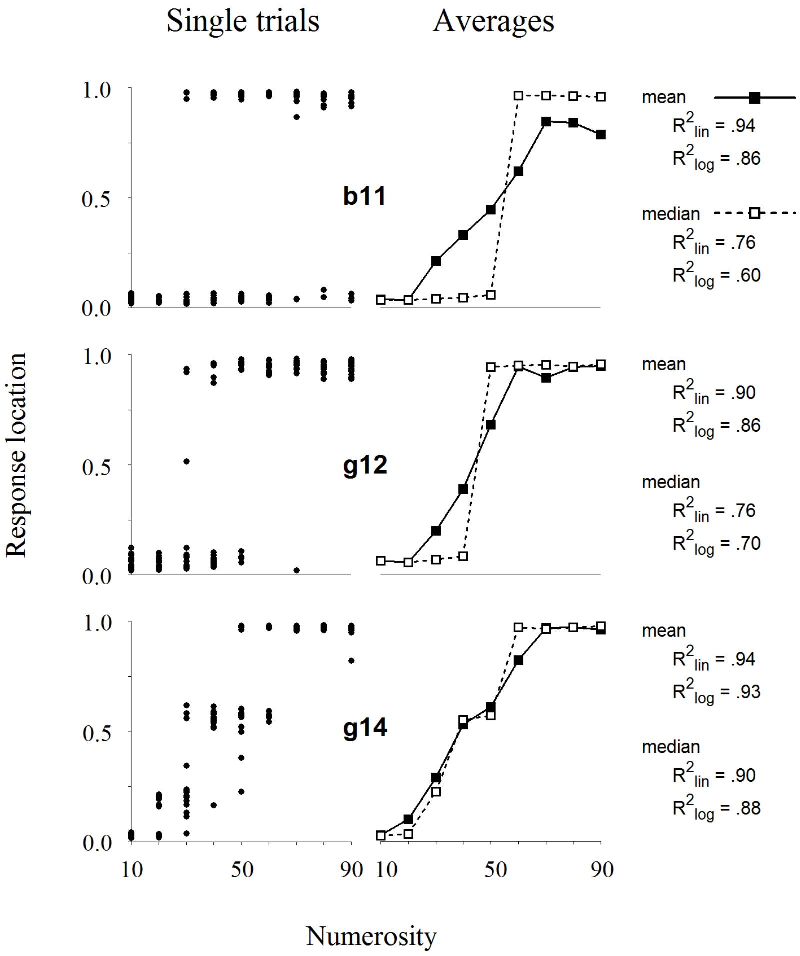


Figure 5. Numerical mapping of three children participants, b11, g12 and g14. For each participant, the left graph depicts single-trial responses and the right graph the mean and median response location curves. The mean is depicted by the filled square data points connected by a solid line, the median by the unfilled square datapoints connected by a dashed line. The far-right side of the figure presents the coefficients of determinations of the best fit linear and logarithmic functions.

Another point pertaining to the discussion about individual average curves, which is also illustrated in Figure 5 by participants b11 and g12, is that many child participants present S-shaped curves, which suggests the occurrence of non-continuous, categorical responding. Such categorical responding entails that in a within-participant comparison between the mean and median measures, the latter should reveal steeper sigmoid-like curves, as is actually shown in our results (refer to Appendix C for the complete individual plots of mean and median average location).

The data, so far, have been consistent with our previous studies. For instance, there are clear inconsistencies when interpreting children's performance across units of analysis. Evidence has been gathering in support of rejecting group level as the sole unit of analysis in Number-to-position procedures. That is to say, children's mean or median group and individual curves have been proved not to describe or "summarize" the basic features of the distribution of responses; rather, we believe that in our inspection of the data thus far, it has been congruent and steadily emphasized the utmost necessity of inspecting single-trial data. As such, this unit of analysis will be the focus in the remaining of the discussion.

3.2.3 Individual analysis - single-trial scatterplots and entropy scores

Adults. We start with adults' data, because their single-trial distributions are well described by the average curves. To better illustrate the overall individual patterns, in Figure 6 we present six exemplars, three subjects from each experimental group (but refer to Appendix D for the scatterplots of the 24 adults).

Almost all adults kept the '10'-leftmost anchor, the exception being participants w16 and w18 from the Control group. Seven participants from each group lost the largest anchor, '90'-rightmost position. The two participants who presented responses farthest away from the trained endpoint are m14 and m6, both depicted in Figure 6.Regarding participant m6, for instance, one observes that though both the data points' "dispersion" and the individual mean increase as a function of numerosity, he practically restricted his responses in the region from the left endpoint up to the midpoint of the bar. To a lesser extent, m14 also preferably responded within the first half of the bar. The remaining participants distributed their responses more evenly along the full extension of the response bar, similarly to what is illustrated in Figure 6 by the remaining four cases.

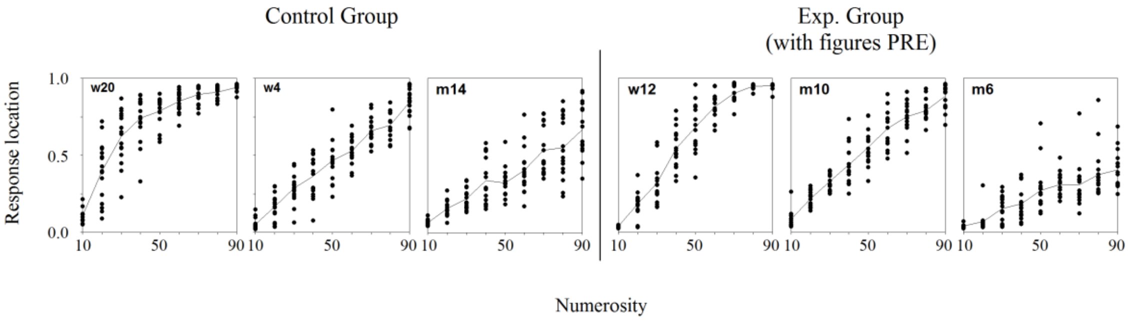


Figure 6. Results of six adult participants: w20, w4 and m14 from the Control Group, and w12, m10 and m6 from the Experimental Group. Each data point is the response location on a single trial; the line is the mean location.

The fitting of individual functions to the average curves hinted that more linear cases (i.e., $R^2_{lin} > R^2_{log}$) have occurred in the Experimental group than in the Control group. Recall, however, that the effect was doubtful or, at best, weak. In order to verify whether it also translated into a more even distribution of single responses along the bar's extent, we computed each participant's normalized entropy score (H).

As illustrated in the right portion of Figure 7, high entropy scores were found in all adult participants. The lowest H value belonged to participant m6 (H = .74), from the Experimental Group, which was to be expected given that he restricted responses to the initial half of the response bar (see Figure 6). The second lowest value was w20's (H = .85), from the Control group, which is also depicted in Figure 6. This value is explained by the "gap" between locations .20 and .40 and by a strong concentration of responses in the second half of the response bar. The remaining entropy score values are all high, at about .92 (refer to Appendix E for the complete individual H scores). An independent samples t-test revealed no significant difference between participants' H scores in the Experimental group (M = .93, SD = .06) and those in the Control group (M = .94, SD = .04) (t(1,22) = 0.43, p = .67).

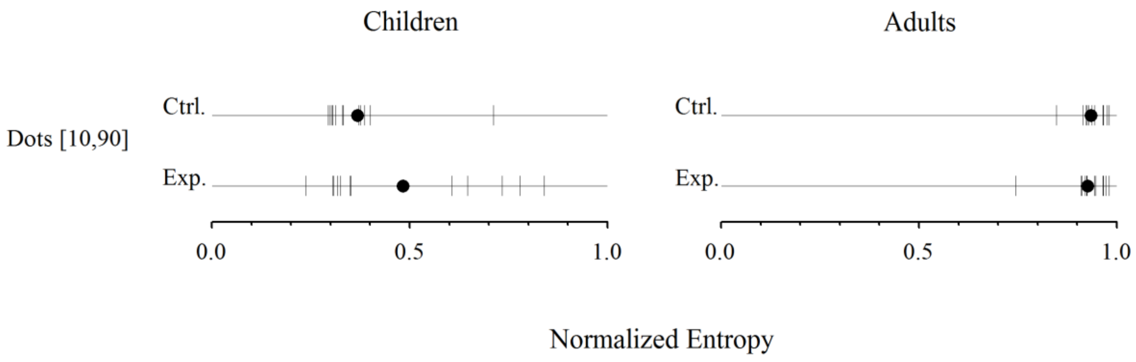


Figure 7. Individual normalized entropy scores (H). H scores are separated by experimental group (thus, n = 12 on each graph). A vertical line represents a participant's H score, and the black dot the group's mean.

Analysis of the H scores, together with visual inspection of single-trial scatter plots, thus confirm that adults' average response location is a good descriptor of their individual performance, since all adults responded in a continuous pattern. Additionally, the experimental manipulation of the pre-training in brightness did not lead to differences between the groups regarding the use of the response bar. Indeed, even when we rank the participants according to their H score, the two experimental groups are equally divided. That is, three members of each experimental group can be found in each quarter of the H scores ranking.

Adults' results are congruent with the previous analyses at the individual and group level curve fittings. First, in regular (control) conditions, all adults seem to already respond along the extent of the response bar. Moreover, as indicated by the nearly ceiling values in the H scores, they tend to respond linearly. In conclusion, having experienced a pre-training which required them to touch five evenly spaced locations on the response bar did not seem to improve an even spacing when adults were required to map numerosities onto the bar.

Children. In Figure 5, we have presented three participants' single trial data points and average curves. The within-subject comparison between the two types of plots illustrated how the average measure could wrongly lead to an interpretation that the child had responded along the bar, in a more logarithmic or linear-like distribution of spatial responses.

As was the case in Study 2, there were three main types of single trial scatterplots: those that presented a bi-categorical distribution, those that presented a tri-categorical distribution and, finally, the cases where more than three location clusters were selected along the response bar. We catalogued each child in terms of his/her distribution of responses along the bar constituting 2, 3, or more than three clusters. This characterization was undertaken by visual inspection of the single-trial scatterplots and by performing a k-means cluster analysis to determine the cut-off point for the number of portions of the bar selected by the participant. In Figure 8, the histogram represents the percentage of these patterns, separated by experimental group.

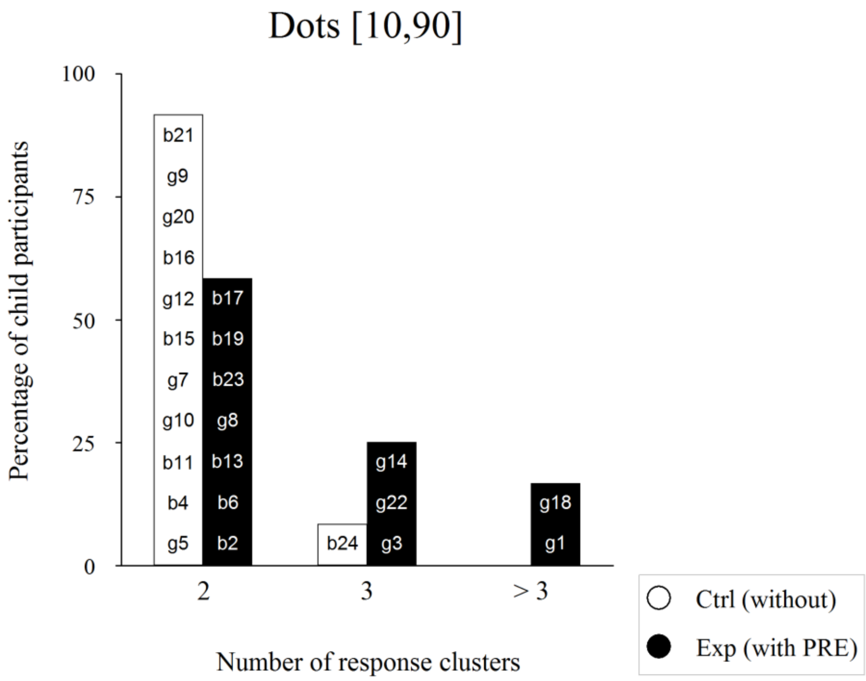


Figure 8. Percentage of children whose response distributions are better described as two, three or more than three response clusters. The white bars represent the Control Group, and the black bars the Experimental Group (with the Figures-to-Position pre-training).

Contrary to adults' performance, the majority of children did not respond continuously, and restricted their responses to two or three sites in the response bar (refer to Appendix D for the scatterplots of the 24 children). Regarding the Control group, whereas b24 restricted his responses to the endpoints plus the midpoint of the response bar (tri-categorical pattern), the remaining eleven participants solely selected the bar's endpoints (bi-categorical pattern). As for the participants in the Experimental group, who had previously been trained to touch differentiated locations of the response bar as a function of images of 5 cartoons, one observes that the number of bi-categorical patterns is smaller than in the Control group. There were two more tri-categorical participants in the Experimental group than in the Control group. In addition, two participants in the Experimental Group (g1 and g18) responded at more than three locations of the response bar. However, a Chi-square test revealed that these differences were only marginally significant ($\chi^2(1) = 3.556$, $p = .059$). It follows that, pertaining to the number of selected clusters, the pre-training tended to improve responding along the bar.

The analysis of the entropy scores is only partially consistent with this weak increment in the use of the bar. On one hand, entropy scores in the Experimental Group ($M = .48$, $SD = .22$) are higher than in the Control group ($M = .37$, $SD = .11$) (see Figure 7 and Appendix F). But on the other hand, this difference fails to reach statistical significance ($t(1,22) = -1.61$, $p = .122$). The statistical significance of group difference aside, it was verified that the lowest H scores belonged to the participants that responded bi-categorically (a moderate Spearman rank correlation coefficient, $\rho\ (23) = .45$, $p = .029$).

Another point of interest was that no child lost the smallest anchor mapping (i.e., 10-leftmost). The three youngest children in the Control group (b4, g5, and b11) lost the largest anchor mapping (90-rightmost). In the Experimental group, only g8 lost the largest anchor. Regardless of experimental group, these all cases presented a bi-categorical pattern. Even so, one could not establish a relationship between preservation of the previously trained anchor mappings and the cluster distribution of single responses. This seems related to the fact that the majority of the children (about 91.6%) restricted their responses to the endpoint positions or the endpoints plus the midpoint.

To conclude, one must emphasize that no child's distribution of single responses could be described as a truly continuous pattern; not even g1, who presents both the largest number of clusters and the largest H score. In other words, in no child's single trial scatterplot does one observe the spread of single response locations accompanying the increment in the mean. The pre-training was not sufficient for children to produce a continuous pattern, similarly to those observed in adult participants. Not surprisingly, then, an independent samples t-test on H scores revealed a significant difference between children's (M = .43, SD = .18) and adults' (M = .93, SD = .05) use of the response bar; $t(1,46) = -13.14$, $p < .001$).

3.3 Counting and Verbal Estimation Assessment

3.3.1 Abstract counting

Recall that the counting assessment was taken solely by the children. The median of the maximum number that children could count up to was "20", with the range going from "2" (b11) to "59" (g9). With the exception of participants b11 and b2 (maximum = "5"), children could count at least up to ten. Only participants g9, g10 and g18 (12.5%)' counting routine surpassed number 30.

3.3.2 Verbal Estimation

Adults. As depicted in the group histogram in Figure 9, all number words between 1 and 9 occurred in equal frequency, hence the relative frequency bars at about 0.11. In fact, all individual H scores equaled 1. Also indicated in the figure by the diagonal line in the left graph, the estimates were perfectly accurate (no errors occurred). The adults did not count the dots aloud, but because the sample presentation time was neither manipulated nor controlled, it is still possible that they have counted subvocally. When they were presented with sets of 10 to 90 dots, the number words within this range were offered less evenly, but nevertheless high individual H scores were obtained (M = .94, SD =.05). As was reported both in Study 2 and by other authors (e.g., Whalen, Gallistel, & Gelman, 1999; Cordes, Gelman, Gallistel, & Whalen, 2001; Lipton & Spelke, 2005), mean estimates increased monotonically with the sample (linear regressions of individual mean estimates yielded R^2 ranging between .85 and .99, with a mean of .95). Regarding the dispersion of the estimates, standard deviations increased monotonically up until "80", then decreased for sample "90".

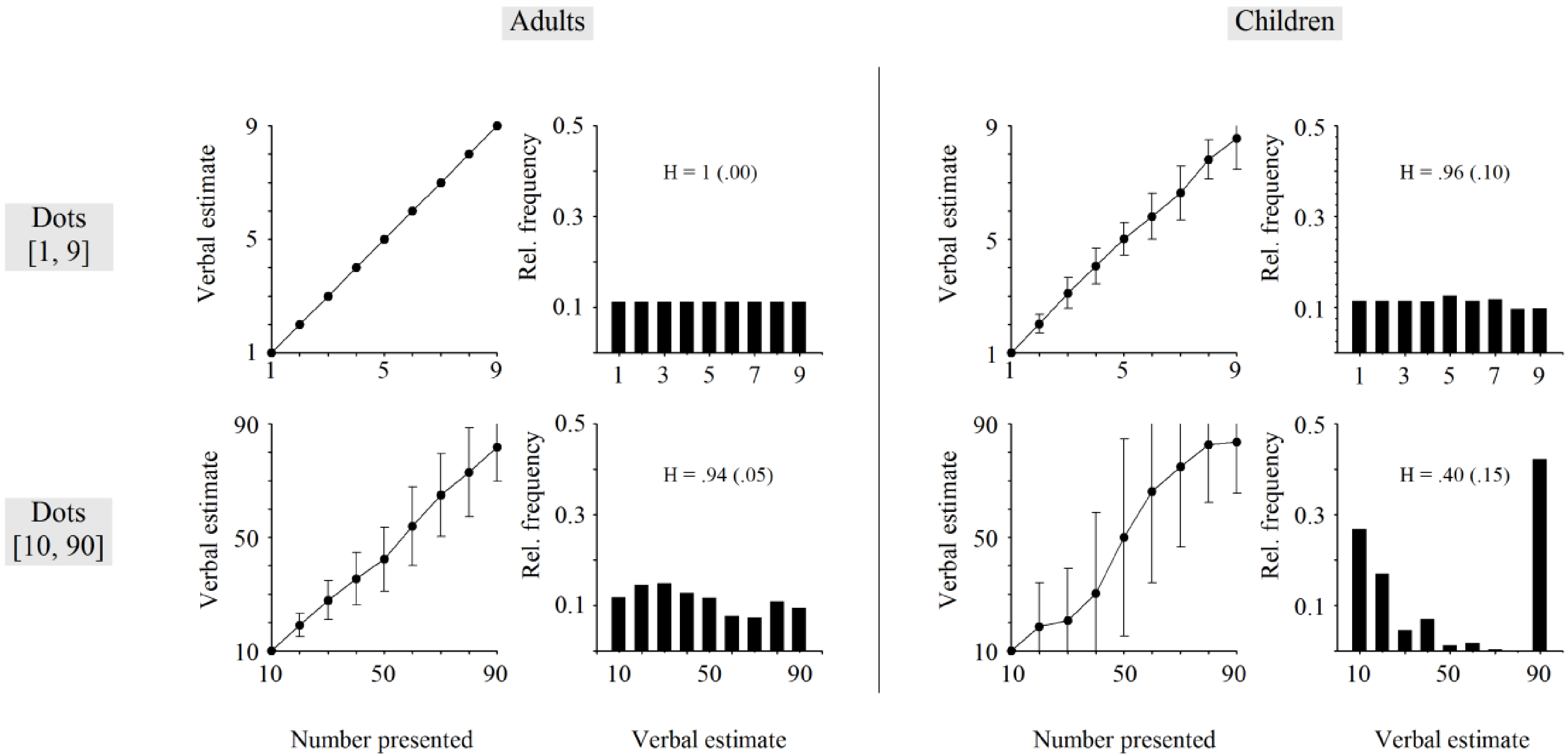


Figure 9. Results of the Verbal Estimation tasks. On each age group section, the left graphs present the group's mean (± 1 SD) verbal estimates (response) as a function of the sample numerosities. The right graphs are the group's histogram of the verbal estimates, and also depict the mean and standard deviation of the individual entropy scores.

As for the coefficient of variation (CV = SD / Mean), it followed an inverted U shape as a function of sample numerosity. The smallest CV value was at anchor “10”, at about .03, then CVs increased up until .27 in sample “50”, and then progressively decreased until the largest anchor, “90”, where CV equaled .15. The same inverted U schape relation between CVs and sample numerosity had been found in our previous verbal estimation results. As we had mentioned in Study 2 when we were discussing why, contrary to some other studies (Whalen et al., 1999; Cordes et al., 2001; Boisvert et al., 2003; Tan & Grace, 2012), our verbal estimation data failed to show scalar variability, we hypothesize that these differences could have derived from procedural features such as providing accurate information about the testing range, as well as feedback to the anchor values.

An additional finding, which had been reported both in Study 2 as well as in other authors’ estimation studies, was that the majority of adults’ verbal estimates were multiples of 10 (see, e.g., Dehaene, Dupoux, & Mehler, 1990; Dehaene & Mehler, 1992; Lipton & Spelke, 2005). This is illustrated in Figure 10, which presents the frequency distribution of the estimates, by the colored series peaking at the decade numbers (x-axis).

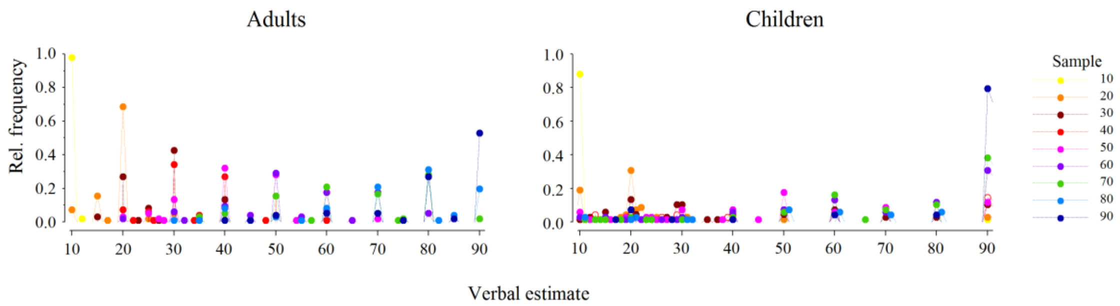


Figure 10. Distribution of the verbal estimates produced in the Dots[10,90] range. The graphs depict the relative frequency of each number word produced, and the colored series discriminate for which sample they were emitted.

In fact, in the current study 88.9% of adults’ estimates were decade numbers. At the end of the session, when questioned informally by the experimenter about their selection of number words, adult participants claimed they could not explain their strategy. When the experimenter pointed out that they had given mostly multiples of ten, but not, e.g., intermediate numbers such as “54”, they were surprised about the possibility that the latter could have been presented. Moreover, they would add that they thought it was more “logical” and “natural” to offer round numbers. Some participants also reported that they gave round numbers so they would increase their chance to give an answer closer to the target. Because the current study replicates our previous results with adults’ estimation, we refer the reader to the discussion in Study 2 on how adults may have responded with verbalizations mediating the perceived (i.e., analog) numerical quantities.

Children. The upper right histogram in Figure 9 shows that children offered the number words between 1 and 9 at approximately the same frequency (individual H scores averaged .96, SD = .10). Also, as their group curve suggests, mean estimates were fairly accurate but the number of errors tended to increase with the sample value. A percentage of 15.1% of children’s estimates were incorrect. Among the errors, 70% differed from the target in ± 1 value. Most errors (76%) consisted in the child offering an estimate smaller than the target (underestimation). Likewise, standard deviation increased with the sample numerosities. For sample “1” it was 0.00 and for “9” it was 1.10. Except for sample “1”, mean CVs tended to decrease with the numerical sample. However, a discussion on how scalar variability holds on preschoolers’ numerical estimation must be put aside because, as had happened in the Study 2, the task was in reality a counting one. That is, the experimenter observed all children counting the numerical sample, especially sets of 4 or more dots. It follows that, in the current study, variability is due not to noise inherent to an analog magnitude representation of numerosities (“number sense” or psychological scale, depending on each author’s preferred term), but to the occurrence of miscounts (skips) and, to a lesser extent, doublecounts errors.

As was the case in the smaller range, children’s mean verbal estimates of the sets composed of 10 to 90 dots increased monotonically with numerosity (see the scatterplot at the second row scatterplot in the right portion of Figure 9). However, the large SD bars immediately accentuate the high level of variability associated with these estimates. The depiction of standard deviation as a function of sample resembles an inverted U shape (graph not shown), with standard deviation increasing from the smallest sample (“10”), where SD = 0.46, up to the midpoint sample “50”, where SD equaled 34.7. SD then subsequently decreased to a new minimum of 18.3, at the largest sample (“90”).

Even more revealing of children's performance than the analysis of the standard variation of their estimates is the analysis of these estimates' frequencies. The rightmost lower histogram of Figure 9 reveals how unevenly were the estimates offered along the [10,90] range. One observes that the most frequent estimates were number words in the bin [81,90] (relative frequency = .42). Next, came the smaller number words, in bin [1,10] (rel. freq. = .27) and bin [11,20] (rel. freq. = .17). Number words in the remaining intervals occurred rarely or not at all. For instance, there was only one estimate that fell within the interval [61,80]. In Figure 10, we progress one step further and discriminate frequency by unit of estimate. This allows us to identify the three main features of children's estimation of large numbers. First, number words larger than "thirty" consisted mostly of decade numbers (amongst the verbal estimates within the interval [31,90], 98% were decade and 2% were non-decade numbers). Second, although decade numbers once again prevailed, in the smaller range [10,30] there were considerably more occurrences of non-decade numbers (30% of the estimates were non-decades).

Lastly, related and complementing the previous point, most children were witnessed trying to count the dots. We noted that children tended to start the trial by consecutively pointing at the dots, one-by-one, while counting them aloud. Initially, after they had individuated all the items children would provide a verbal response (number word). But for most of the session, the events progressed differently. For instance, in certain trials the maximum time elapsed and the trial was repeated. After the ITI, the child either continued counting the remaining, still un-pointed dots or would give up the counting strategy. When the latter happened, the estimate was either the last uttered word in the counting series, a "novel" number word or, especially when the sample was larger, the largest anchor word, "ninety" (about whom the child had been instructed). Thus the high frequency of "ninety" responses, as observed in Figure 10, following a wide range of numerical samples but the three larger values in particular (note the green and blue series peaking at x-axis value 90).

In our Study 2 we found the same overall results regarding children's estimates of larger numbers. What is more, in the current study we corroborate that an average estimation curve, particularly a group curve, does not represent actual individual behavior. Instead, children's performance hinted at the play of different strategies when the participants are asked to estimate smaller or larger sets of dots. On one hand, it seems that counting was the basis for responding in the Verbal Estimation Dots[1,9] task. On the other hand, their behavior towards sets of [10,90] dots was not as clear-cut.

Performance was possibly a joint result of (a) tagging elements one-by-one (counting), (b) estimating the remaining, to-be-counted dots, and (c) a bias to offer the largest anchor word ("ninety").

Although beyond the scope of the current experiment, we believe it would be interesting to manipulate procedural features such as the sample duration (e.g., Beckmann, 1924, cit. by Gelman & Gallistel, 1978, p. 69; Gelman , 1972; Gelman & Tucker, 1975) and the amount of training and feedback (e.g., Gelman, 1969), and determine under which conditions the preschoolers' response strategies balance counting - with accurate representations of the arrays -, and estimation - yielding only approximations (Klahr & Wallace, 1973).

3.4 Relation between number-to-position estimates, counting and verbal estimates

To investigate the relationship between number-to-position performance and counting or estimation abilities in preschool children, we conducted Spearman's rank order correlation analyses between these variables. Each child was classified in age (months), pre-training experienced (CtrlGroup – 0, ExpGroup – 1), use of response bar during the Number-to-Position task (entropy score: H_{90}), an Abstract Counting score (largest number word in counting series) and a measure of the verbal estimates offered during the Verbal Estimation tasks (entropy scores: H_{V-9}, H_{V-90}). Table 3 presents the Spearman's rank order correlation coefficients obtained.

First, age was found to correlate moderately with counting proficiency. Not surprisingly, our data converges with the literature in that the older the child, the higher he/she can count up to (e.g., Fuson, 1988). Age was also moderately associated with the verbal estimates in the [1,9] range (H_{V-9}). This is probably due to the fact that the estimation task was addressed by children as a counting task. In line with this hypothesis, counting was found to strongly correlate with H_{V-9}. However, perhaps because the range in the Verbal estimation of Dots[10,90] task was well beyond any child's counting word series, a similar relationship between age, counting and estimation performance was not verified with larger numerosities (i.e., the [10,90] range). In the previous section, we have already described that children exhibited different strategies when asked to estimate smaller or larger sets of dots. For instance, sometimes counting progressed until the limit of the child's word series, but even more frequently, *tagging* was interrupted before that and the child responded “ninety”. This overall bias to treat sets of 40 or more dots as undifferentiated “nineties” may be have been shown by good and bad counters alike, which in turn would explain why the moderate correlation between the two verbal estimation tasks (H_{V-9} and H_{V-90}) did not reach statistical significance.

Table3

Spearman correlations between children’s performance in the number-to-position, the counting and the verbal estimation tasks

Variable	PRE-train	Age	H_{-90}	CtAbs	H_{V-9}
PRE-train	–				
Age	-.02	–			
H_{-90}	.21	-.11	–		
CtAbs	-.18	.51*	-.15	–	
H_{V-9}	-.04	.46*	.19	.61**	–
H_{V-90}	.04	.17	-.11	.28	.35

Note. PRE-train: absence or presence of a pre-training {0,1}; H_{-90}: entropy scores of the spatial responses during the number-to-position Dots[10,90] task; CtAbs: score in the Abstract Counting task; H_{V-9} and H_{V-90}: entropy scores of the verbal estimates.
* $p < .05$ ** $p < .01$

Earlier in the text we discussed the, at most, limited effect that the Figures-to-position pre-training exerted in the number-to-position task. In Table 3, one observes that neither counting nor estimation performance were found to be significantly correlated with performance in the number-to-position task, as measured by the entropy of the distribution of spatial responses across the response bar (H scores).

Lastly, how do counting and estimation proficiency relate to performance in the number-to-position task? Lipton and Spelke (2005)’s hypothesis, shared with other authors (e.g., Berteletti et al., 2010), is that familiarity with numbers is associated with performance in number-to-position tasks. However, when we measured this performance in terms of use of response bar (i.e., entropy scores, H_{-90}), no significant correlation was found. Granted, the authors who advocate a log-to-lin shift in the mental number line could counter that linearity would be a better metric. For that reason, and despite how poorly children’s individual performance is characterized by a simple linear model, we repeated the correlations but now the index for performance was the variance accounted for by the best fitting linear model (R^2_{lin}), both in the number-to-position and the two estimation tasks. Table 4 summarizes the results.

Table 4

Spearman correlations between children's linearity in the number-to-position task and performance in the counting and the verbal estimation tasks

Variable	PRE-train	Age	R^2_{-90}	CtAbs	R^2_{V-9}
PRE-train	–				
Age	-.02	–			
R^2_{-90}	-.01	-.12	–		
CtAbs	-.18	.51*	-.25	–	
R^2_{V-9}	-.03	.35	.20	.61**	–
R^2_{V-90}	-.19	-.09	-.16	.17	.13

Note. PRE-train: absence or presence of a pre-training {0,1}; R^2_{-90}: R^2 values from the best fitting linear function of number-to-position Dots[10,90] task; CtAbs: score in the Abstract Counting task; R^2_{V-9} and R^2_{V-90}: R^2 values from the best fitting lineal model of verbal estimates.

* $p < .05$ ** $p < .01$

Most important, the coefficients of determination in the number-to-position task (R^2_{-90}) were not significantly correlated with either of the verbal estimation tasks (R^2_{V-9} and R^2_{V-90}) or to the counting proficiency. In other words, linearity in the number-to-position Dot[10,90] task could not be predicted by performance in the verbal estimation or abstract counting tasks. In conclusion, we found no evidence for a relation between children's familiarity with numbers, counting or estimation abilities and their spatial mapping of non-symbolic numerosities onto space. The same conclusion was reached at the end of our Study 2.

4. Conclusions

In the current study, a Figures-to-Position pre-training condition was set up to test for the effect of having responded once on a spatial *continuum*, when preschool children and adults are tested in a Number-to-Position task. The main difference with the pre-training we had implemented before, the Brightness-to-position pre-training and the current one, was that in the current pre-training the samples were not ordered along a physical *continuum*.

Eleven of the twelve preschoolers in the Control group restricted their responses to the anchor positions, exhibiting what we refer to as a bi-categorical pattern. In contrast, the number of bi-categorical subjects in the Experimental group, which experienced the Figures-to-position pre-training, decreased to seven. In addition, although normalized entropy (H) scores were higher in the Experimental group than in the Control group, these differences were not statistically significant. At best, we may interpret these results as the Figures-to-Position pre-training having a weak effect on having participants select locations other than the two anchors, in the subsequent Number-to-Position task.

Related to this issue of the prevalence of categorical responding in preschoolers, on the total of the 24 children tested, only 2 participants, both from the Experimental group, distributed their responses at more than three clusters. But what matters most, is that no child's distribution of single responses could be described as a truly continuous pattern, as was the case in all adult participants, Indeed, we believe that a major finding in the current study was the replication of our previous studies' data with preschoolers. To specify, again we observed that adults' average response location was a good descriptor of their individual performance (they used the extent of the response bar). Additionally, preschoolers' mean or median group and individual curves proved not to be able to describe or "summarize" the basic features of the distribution of single-trial responses (Speelman & McGann, 2013; Trafimow, 2014).

CHAPTER V : STUDY 4

Control of a response continuum by the numerical stimulus continuum: isolating the effects of a perceptual training on Number-to-Position performance

1. Introduction

In the preceding study we have singled the effect of a "manual" component on Number-to-position performance. Namely, we pre-trained participants to touch the response bar (spatial *medium*) as a function of cartoon TV characters. In this Figures-to-position pre-training the stimuli were not ordered along a continuum. We had hypothesized that the preschoolers exposed only to the Number-to-Position conditions did not respond along the bar because they had never responded along ordered spatial positions. We conjectured that, the verbal instruction to do so when they are tested with numerosities could have been made clearer to the participants, had they ever touched the response bar at locations other than the endpoints. However, our procedure did not improve significantly the use of the response bar.

In the current study we continue to investigate the variables that may affect the occurrence of continuous patterns of responding in a Number-to-position task. However, we will move from the previous' study enquiry pertaining to response topography features, to features of the sample stimulus *continuum*: numerosity. Specifically, we will investigate whether a perceptual training on the numerical samples may affect how these samples are mapped onto space.

Experimenter's notes from our previous studies lead us to expect that some preschoolers may have had difficulties discriminating among the dot arrays in the [10,90] range. We observed that, when tested in the Numbers-to-position task, some preschoolers would offer spontaneous verbalizations such as "many", "the most" when they were presented a numerical array. Such particular verbalizations accompanied the selection of the response bar's rightmost endpoint. Most important, they seemed to occur when the numerical samples were arrays from 40 dots onwards. In a subsequent experimental phase, preschoolers were tested with the same numerosities, but asked to offer a verbal estimate. We found that, when the arrays were bigger than 30 dots, the majority of the estimates corresponded to the largest anchor number ("ninety"). These observations lead us to question whether some children had difficulties perceiving the magnitude differences within the tested numerical range.

These findings may be related to a phenomenon which has been reported very early in the field of infant numerical cognition. In 1921, psychologist A. Descoeudres conducted a series of tasks in which about 3 and a half-years old children had to quantify sets of items (Descoeudres, 1921, as cited in Gelman & Gallistel, 1978, p. 54).

Children's performance for numbers one, two and three was highly accurate, across a variety of tasks. When the targeted number was four or more, performance dropped sharply. Notably, the study also included descriptions of children's verbalizations and other observation measures. These revealed that children treated sets larger than four as undifferentiated and all equal to "a lot". Descoeudres called this drop in accuracy the "*un, deux, trois, beaucoup*" phenomenon ("one, two, three, many"). The "*un, deux, trois, beaucoup*" phenomenon raised the question whether children could not differentiate large sets from each other. Several decades later, Gelman and Tucker (1975) presented 3- to 5-years old children tasks similar to those in Decouesdres' study, but in addition to manipulating the number of elements in the array (2, 3, 4, 5, 7, 11, 19), they manipulated also exposure time to the numerical array (1, 5, or 60 s) and whether each array was composed by homogeneous or heterogeneous elements. The authors found that the younger children showed the "*un, deux, trois, beaucoup*" phenomenon, but, and of particular interest to us, there was a rank order correspondence between their verbal estimates and the numerosities (Gelman, 1977; Gelman & Gallistel, 1978, p.56). In other words, although accuracy was low for sets larger than 5, even in these cases the larger the array, the larger the offered verbal tags (number words). This indicated that "children are sensitive to the ordinal characteristics of larger numerosities, to the ordinal characteristics of the number-word sequence and to the conventional relation between the two" (Gelman & Gallistel, 1978, p. 62).

Perhaps categorical performance in the Number-to-position task is also a type of "*un, deux, trois, beaucoup*" phenomenon. According to this supposition, if a child does not perceive the magnitude differences among the larger sets of dots, she responds at the largest anchor position. On the other hand, during previous Number-to-position and Verbal Estimation sessions some children occasionally provided imprecise verbal quantifiers such as "*very few*", "*some*", "*more or less*", "*a lot*", "*many*", or "*the most of all*". In the case of Verbal Estimation tasks, because participants were demanded to answer with a number word, sets larger than 30 tended to be treated as undifferentiated "*beaucoups*", whose label was the largest number word that was communicated to them by the experimenter ("ninety"). However, our notes suggest that there is a difference between the two estimation tasks (verbal vs. spatial) regarding the occurrence of spontaneous verbalizations and of counting behavior during the presentation of the numerical arrays. Counting is mostly absent during the Number-to-position task and even when it occurs, it stops after a few trials. Perhaps the absence of a specific tag (in this case, a number word) leads to the numerical stimuli being perceived as undifferentiated “many”. Another possibility is that the creation of certain rules of response or, to put it differently, the mediation of verbal behavior is task-specific and not related to perception.

The experimental manipulation that we introduced in the current study derives from the field of perceptual learning. In the words of Eleanor J. Gibson (1969), perceptual learning is a process by which there is “an increase in the ability to extract information from the environment, as a result of experience and practice with stimulation coming from it” (p. 3). Among perceptual learning studies, those on categorical perception of visual stimuli are of particular interest (e.g., Goldstone, 1994b, 1998; Goldstone, Lippa, & Shiffrin, 2001; Golstone & Hendrickson, 2010; Ahissar, Laiwand, & Hochstein, 2001). Similar treatments to the one we intend to implement were used with size, hue, saturation and brightness discrimination (Burns & Shepp, 1998; Ozgen & Davies, 2002; Roberson, Davidoff, & Davies, 2005; Winamer et al., 2007).

For instance, in Goldstone (1994b)’s study, adult participants were taught to categorize square figures that could vary in size or brightness. A quarter of the participants were trained with four values of size, another quarter with four values of brightness, and another quarter with brightness and size. The remaining subjects were a control group and did not undergo any categorization training. The participants had to press the keyboard-keys 1, 2, 3, and 4, according to the stimulus values on the relevant dimension. Consider the case of a participant in the brightness-relevant categorization training. Whenever the brightest square was displayed on screen, the subject had to select key 1, regardless of the square size. In a subsequent experimental phase, participants entered a same/different judgment task where on each trial two squares appeared on screen, and they were either adjacent exemplars or the same square repeated. Participants had to decide whether the squares were exactly identical on both their size and brightness or if they differed even slightly on any dimension. The authors found that same/different judgments were more accurate in the dimension that had been relevant for categorization. Moreover, the largest difference in accuracy between experimental and control groups occurred with the particular stimulus values that had been trained.

In the current study we will implement perceptual training on numerical discrimination and measure its effects on a Number-to-position task, in comparison with a control condition where only number-to-position is tested. Our hypothesis is that a continuous performance on the Number-to-position task could be enhanced by having the child become more proficient at discriminating the numerical quantities. Our training will require participants to learn to associate five particular numerosities with the images of 5 cartoon TV characters. Rather than testing all the tens numerosities within the [10,90] range, we selected sets of 10, 30, 50, 70 and 90. This selection of five categories of judgment is recommended by early findings on perceptual learning. For example, in Pollack (1952)'s study participants were tasked to assign, by crescent order of magnitude, a number word to *m* tones varying in frequency (e.g., two distinct Hz tones, described with numbers '1' and '2'). Across different conditions, the number *m* of different tones (and rating numbers) was manipulated from 2 to 15. It was found that after 5 tags, performance became alike. To put it in another way, for error-free identification, no more than five alternative stimuli should be used. The same type of protocol has been applied to visual sensory attributes (area of squares: Eriksen & Hake, 1955a, 1955b; length of lines: Baird et al., 1970; visual position: Hake & Garner, 1951), yielding similar conclusions. Because in our study we intend to expedite accuracy in our perceptual training protocol, the training will consist of five numerosity-figure assignments.

We hypothesize that if we provide five different visual tags to the numerosities within the testing range, perhaps this categorization of the numerosities will increase perceptual sensitivity which, in turn, may lead preschoolers to avoid placing their responses at the same location in the response bar. Because the numbers will become more "distinct from each other" and will also be attributed different tags (different figures), children will better understand that different numerosities cannot possibly belong at the same position along the path.

2. Method

2.1 Participants

Twenty four Portuguese pre-schoolers (12 girls) and twenty four young adults (12 women) participated in the study. The mean age of preschoolers was 4.98 years-old (SD = 0.47; range 4.24 - 5.78). The mean age of adults was 20.41 years (SD = 3.51; range 18.32 - 27.24). Informed consent was given by all adult participants and by the children's parents.

2.2 Numerosity stimuli

The numerical stimuli were as described in Study 2.

2.3 Procedure

As in the previous experiments, participants were assigned to one of two experimental groups by first matching pairs in terms of age and sex. Participants were seated in front of a touchscreen laptop, in a separate room of the school. This was the same computer used during the previous studies. The experimenter remained in the room, seated about 0.75 m behind the participant to keep out of his sight and prevent response bias. A separate monitor, positioned behind the participant and facing the experimenter, was connected to the laptop and displayed the experimental events. The previous experimental program in Visual Basic language was used to control all session events and record participants' responses in number-to-position sessions. A new experimental program, also in Visual Basic, was written for the sessions of the Treatment phase.

The current used a pretest-posttest research design, with two experimental groups (Control and Experimental). The experimental conditions and, consequently, the number of experimental sessions, depended on the experimental group the participant was assigned to (Figure 1). Participants from both groups were pre-tested on a Number-to-position procedure with sets of 10 to 90 dots and later post-tested in the same procedure after the experimental treatment was administered to the Experimental Group. Thus, those in the Control Group were solely tested in number-to-position tasks.

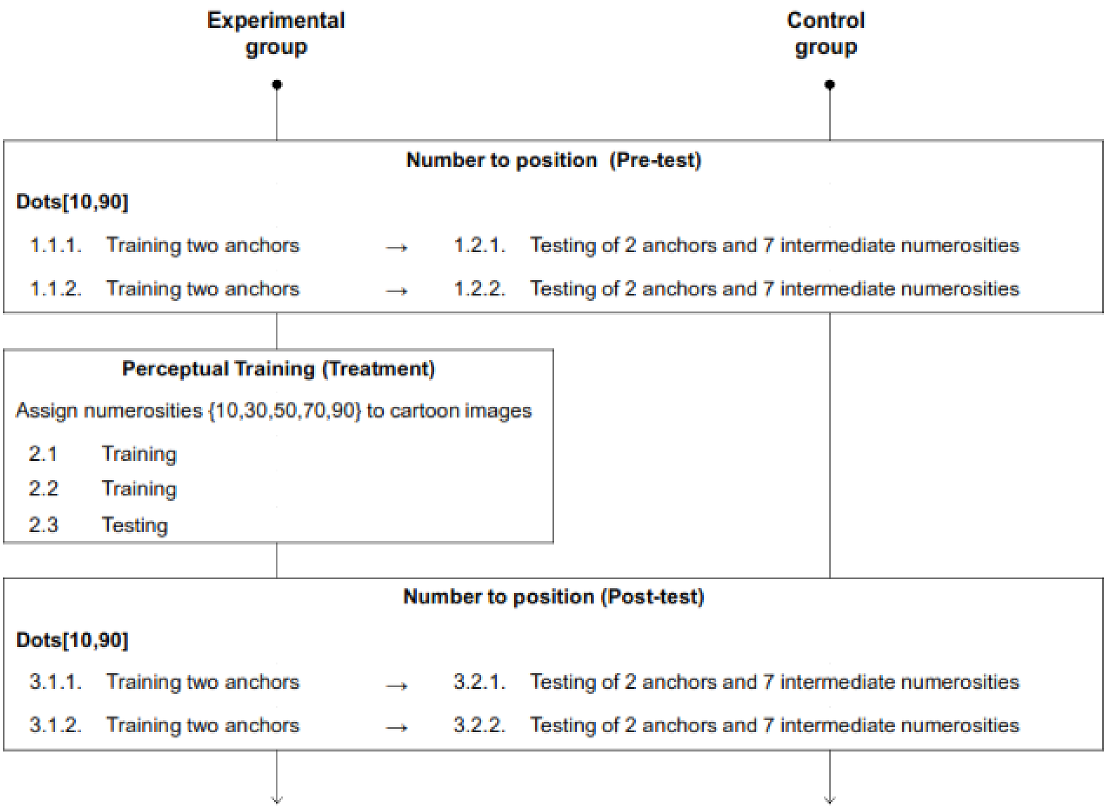


Figure 1. Diagram depicting the progress throughout the experimental phases for each group. Participants in the Experimental group received a Perceptual Training protocol in-between the two Number-to-position assessments.

Pre- and Post-Test in Number-to-Position. The Number-to-position task with arrays of Dots[10,90] remained unchanged from Study 2. To summarize, participants were initially presented the yellow response bar and were told that it was a path. The experimenter told the child where the path started, pointing at the leftmost position. Then the path continued and one could walk and walk up until the path's end (the experimenter moved her finger along the path, in the left-to-right orientation until reaching the rightmost position). The child was then invited to show the experimenter where the path started, and how to move along the path until its endpoint.

The experimenter told the child that the path was where the numbers belonged. In other words, numbers stayed along the path. At the path's beginning (experimenter pointed to the leftmost position) was where the fewest number of dots should be put, and as the number of dots increases, one must touch the path more and more in the direction of its right endpoint. The larger the number, the more it would have to walk along the path. The largest number of dots should be placed at the right endpoint (the experimenter touched the bar at about five positions, ordered from left to right and with constant spacing in-between).

Next, the participants were trained to respond at the anchor numerosities. They were told that they would start by learning which dots belonged to the beginning and end of the path. They started the computerized sessions, and each response was followed by verbal feedback from the experimenter. Additionally, the two-step correction procedure was in effect.

After learning to respond at the anchors, participants were told they would now see not only the smallest and largest possible numbers, but also other numbers of dots. The experimenter continued by reminding the participant that the fewest number of dots belonged at the leftmost position (while pointing there), and the most number of dots belonged at the rightmost position (while pointing there), and they would have to decide where to place the other numbers. The experimenter remarked that the larger the number of dots, the further its location along the path should be (while saying so, the experimenter moved her finger along the path, left-to-right, at five locations). Then the children started the computerized session, without feedback from the experimenter and without the correction procedure.

Perceptual training (Treatment). The treatment sessions were administered only to participants in the Experimental Group (Figure 1). Participants were shown five cards with printed arrays of 10, 30, 50, 70 and 90 dots. Then they were asked to order the cards, from the fewest to the largest number of dots. After they succeeded in ordering the arrays, the experimenter showed the child other five cards with printed images of cartoon characters. The experimenter would place 'Shaun the sheep' above the card with 10 dots and say that when they saw the smallest possible number of dots (while pointing at the card), they had to select the image of 'Shaun the sheep'. Similar instructions were given for the remaining number-figure associations, which were: '30-Donald', '50-Noddy', '70-Winnie', and '90-Tinkerbell'. Next, the child started the training with the computerized procedure.

Each trial began with an ITI of 1.5 s, after which the star image appeared at a random location on the screen. A touch to the star image displayed the numerical stimulus. A set of 10, 30, 50, 70 or 90 dots appeared horizontally aligned and below the upmost part of the screen, just as was the case when it was a numerical-to-position session. Yet, instead of a yellow response bar, at the lower portion of the screen five picture boxes appeared horizontally aligned, with constant spacing amongst them. Each picture box depicted one of the five cartoon characters. The assignment of the figures to the picture boxes was randomly determined across trials. Because on each trial, location of "Shaun the sheep" and the other images could appear in any of the five picture boxes, the child could not respond based on the position

The child selected one of the images by touching it. A touch to an image was signaled by a yellow inverted triangle which appeared above the picture box. The experimenter remained next to the child throughout the session, and would give accurate verbal feedback following correct and incorrect responses. In case of an incorrect response, the two-step correction procedure ensued (as in previous experiments, it was a two-trial repetition, firstly experimenter-guided, and then self-guided).

In each training session, there were eight presentations of each numerical sample, which resulted in 40 trials. The child received two sessions of training, with an interval of about half a day between them. Afterwards, they underwent a testing session during which there was neither feedback nor correction procedure. Following each response, the selected picture box was signaled with the inverted triangle during 750 ms after which a new trial began. During this testing session participants were presented 16 exemplars of each numerical sample, which amounted to (16 * 5) 80 trials.

3. Results and discussion

3.1 Perceptual Training (Treatment)

All participants in the Experimental groups completed the training and testing sessions. The results of the testing session are summarized in Figure 2, where the colored series represent how often each of the five presented images were selected by the participants, after a specific sample (numerosity) (refer to Appendix A for the complete individual results). Also, to inspect differences in accuracy, a mixed between-within ANOVA, with a Greenhouse-Geisser correction, was carried with age group (Children and Adults) as the between factor and numerical sample (10, 30, …, 90) as the within factor.

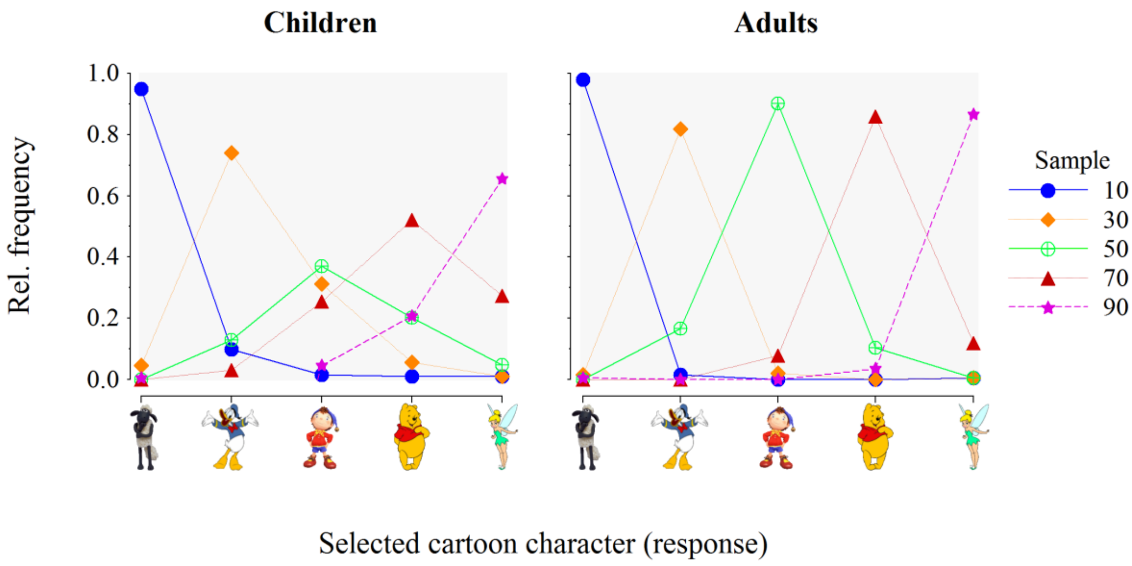


Figure 2. Relative frequency of the images selected during the 'Perceptual Learning' treatment. The colored data point series discriminate following which sample (numerosity) were the images selected, given that the correct Number-Image assignments were 10-Shaun, 30-Donald, 50-Noddy, 70-Winnie, and 90-Tinkerbell.

In Figure 2, each colored series peaks at the correct image, but there are clearly differences between children and adults. The right panel in Figure 2, depicting adults' data, presents high accuracy, given that each series peaks at the correct image with frequency values above .80. Adults' mean percentage of correct responses was 88.4%. The few errors were responses at the immediate vicinity of the correct stimulus. In other words, when adults erred, they tended to select images associated with the numerosity immediately above or below the trials' numerical sample.

Pairwise comparisons revealed that adults' accuracy was significantly higher for the smaller stimulus, '10', compared with the other samples ($p < .05$); the exception was sample '50' where the difference was marginally significant ($p = .069$). The remaining paired differences did not reach statistical significance ($p > .05$).

Not surprisingly, the mixed ANOVA revealed a significant main effect of age group ($F(1,64.913) = 43.466$, $p < .001$), thus confirming that preschool children performed worse than adults. The percentage of children's correct responses averaged 64.7%. Except for the smaller numerical sample (10), correct responses peaked at frequencies below .80. Children's accuracy as a function of numerical sample followed an inverted-U shape, with the poorer performance occurring at the middle sample '50' These results were confirmed when inspecting the paired differences in accuracy between the numerical samples, which revealed significant differences ($p < .05$) between all pairs, except for a not significant difference between '30' and '90' ($p = .740$) and a marginal difference between '70' and '90' ($p = .060$). However, each colored series still peaked at its corresponding correct figure, resembling the typical shape of stimulus generalization gradients (Guttman & Kalish, 1956; Shepard, 1958b; Blough, 1975; Honig & Urcuioli, 1981; Ghirlanda & Enquist, 2003). Indeed, errors were positively associated with the distance between the numerical samples associated with the selected images.

In conclusion, both with children and adult participants, our training protocol seemed to be effective in establishing stimulus control by the numerical dimension.

3.2 Number-to-Position performance

By learning to select a specific image in the presence of a certain number of dots (the “perceptual training” treatment), children were able to discriminate numerosities along the [10,90] range. In other words, during the “perceptual learning” training, large arrays were not treated as undifferentiated “many”, similarly to the ““*un, deus, troix, beaucoup*” phenomenon (Descoeudres, 1921; Gelman & Tucker, 1975). Our hypothesis was that this assignment of a single visual *tag* (i.e., cartoon image) to each of the quantities would make them so “distinctive” that when children were subsequently asked to place these numerosities along a spatial *continuum*, they would avoid clustering them at the same locations.

To answer the question of whether the perceptual training affected number-to-position performance, we compare the pre- and post-test moments. We begin with the mean and median curves from the Control and Experimental groups, which are depicted in Figure 3.

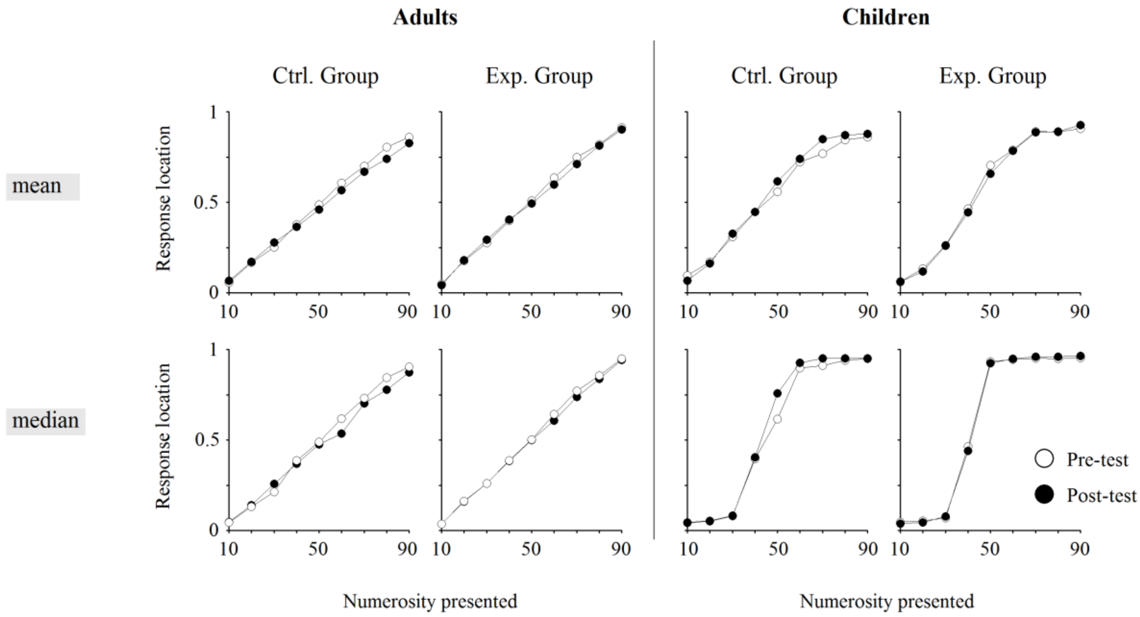


Figure 3. Group response location curves as a function of numerosity, in the Pre- and Post-test Number-to-Position tasks. White dots represent Pre-test performance, and the black dots the Post-test.

The most prominent feature in Figure 3 is that, for both adults and children, pre- and post- curves practically superimpose. This absence of differences between the two test moments is observed either when one attends to mean or to the median response location measure.

As with our previous studies, the distinction between the two average measures is not informative in the case of adults' data, but is so in the case of preschoolers (see Figure 3). In fact, in both children's Experimental and Control group, the mean curves are smooth, with response location gradually increasing as a function of numerosity. In comparison, the median curves are more markedly S-shaped, suggesting that they result from categorical patterns at the individual level of analysis. This is confirmed when inspecting children's single trial scatterplots (see Appendix B).

Since group measures were ill-descriptors of individual performance, the effect of the perceptual training treatment was next inspected by comparing pre- and post-test individual normalized entropy scores (H), goodness of individual linear and logarithmic fitting (R^2_{lin} and R^2_{log}), and, in the case of children participants, the number of response clusters along the bar (k). Adults and children's results are depicted in Tables 1 and 2, respectively.

We first address adults' data. A mixed between-within ANOVA, with experimental group (Control and Experimental) as the between factor and test moment (pre- and post-test) and as the within factor, was used for testing the statistical significance of these variables on adults' use of the response bar (normalized entropy - H - scores). The ANOVA revealed that the effect of test moment did not reach statistical significance ($F(1,22) = .061$, $p = .808$). In other words, adults distributed their responses evenly along the response bar, as indicated by the high H scores (see Table 1), without significant differences between the pre-test and post-test moments. In addition, there was neither a significant effect of experimental group ($F(1,22) = 4.014$, $p = .058$), nor an interaction effect between group and test moment ($F(1,22) = .148$, $p = .704$).

Not only did the perceptual training failed to increase the normalized entropy of responses, a mixed ANOVA with R^2_{lin} as the dependent variable, showed that it also did not increase linearity of responses (no main effect of test moment, $F(1,22) = .044$, $p = .835$, nor of experimental group, $F(1,22)= .065$, $p = .801$, and finally, neither an interaction between these two factors, $F(1,22) = .114$, $p = .739$)).

Table 1

Adults' Pre- vs. Post-test Number-to-Position performance.

		Pre-			Post-		
		H	R^2		H	R^2	
			Lin	Log		Lin	Log
Ctrl.	**m3**	.92	.94	.94	.93	.98	.97
	w10	.91	.99	.90	.85	.97	.87
	m11	.99	.99	.95	.98	.99	.94
	w12	.95	.98	.88	.97	.99	.88
	w13	.95	.98	.85	.89	.99	.92
	w14	.96	.99	.90	.98	.99	.92
	m15	.91	.98	.83	.93	.99	.86
	m16	.95	.97	.96	.96	.98	.96
	m17	.94	.98	.91	.81	.95	.95
	w21	.71	.98	.93	.93	.97	.91
	m22	.94	.97	.80	.89	.95	.75
	w23	.97	.96	.98	.96	.95	.97
	Avg	.92	.98	.90	.92	.98	.91
	SD	.07	.02	.05	.05	.02	.06
	Group		1.0	.92		1.0	.93
Exp.	**w1**	.99	.99	.92	.99	1.0	.90
	m2	.92	.98	.83	.93	.96	.80
	m4	.92	.96	.98	.90	.87	.99
	w5	.96	.99	.95	.98	.98	.96
	w6	.95	.98	.96	.94	.99	.93
	m7	.97	.96	.97	.94	.96	.99
	w8	.95	.97	.86	.96	.99	.89
	m9	.88	.96	.90	.95	.99	.90
	w18	.97	.99	.87	.99	.99	.90
	w19	.94	.99	.92	.93	.99	.94
	m20	.98	.99	.91	.99	1.0	.90
	m24	.99	.99	.94	1.0	1.0	.91
	Avg	.95	.98	.92	.96	.98	.92
	SD	.03	.01	.05	.03	.04	.05
	Group		1.0	.93		1.0	.93

Note. (H): normalized entropy scores; (R^2): coefficients of determination yielded by the fitting of simple linear and logarithmic functions to each participant's mean response location and to the group mean curve.

Table 2

Children's Pre- vs. Post-test Number-to-Position performance.

		Pre-test				Post-test			
		H	R^2		k	H	R^2		k
			Lin	Log			Lin	Log	
Ctrl.	**g17**	.59	.95	.84	3	.62	.97	.89	3
	b2	.45	.58	.83	2	.22	.52	.79	2
	b9	.29	.87	.65	2	.36	.89	.77	2
	g11	.66	.95	.88	2	.38	.86	.67	2
	g10	.60	.90	.84	3	.60	.92	.81	3
	b4	.48	.93	.93	3	.43	.89	.84	2
	g14	.96	.97	.87	>3	.86	.95	.78	>3
	b7	.73	.92	.78	>3	.41	.79	.88	2
	g24	.62	.92	.83	3	.44	.91	.91	3
	b16	.68	.82	.94	2	.76	.72	.90	2
	g22	.85	.93	.83	>3	.78	.96	.94	>3
	b18	.55	.80	.87	2	.40	.79	.91	2
	Avg	.62	.88	.84		.52	.85	.84	
	SD	.18	.11	.07		.20	.13	.08	
	Group		.97	.94			.96	.94	
Exp.	**g5**	.38	.66	.88	2	.35	.81	.89	2
	b1	.31	.71	.85	2	.31	.87	.93	2
	b12	.40	.87	.82	3	.31	.85	.80	3
	b3	.68	.97	.83	>3	.72	.95	.86	>3
	g13	.31	.90	.85	2	.33	.78	.64	2
	g6	.35	.87	.81	2	.38	.83	.85	2
	b8	.33	.92	.74	2	.34	.86	.64	2
	g20	.48	.88	.94	3	.49	.84	.97	3
	g21	.88	.98	.91	>3	.85	.92	.77	>3
	b19	.50	.90	.92	3	.55	.91	.96	3
	b15	.69	.93	.97	>3	.90	.99	.91	>3
	g23	.32	.76	.81	2	.36	.72	.83	2
	Avg	.47	.86	.86		.49	.86	.84	
	SD	.19	.10	.07		.22	.07	.11	
	Group		.93	.92			.95	.92	

Note. (H): normalized entropy scores; (R^2): coefficients of determination yielded by the fitting of simple linear and logarithmic functions to each participant's mean response location and to the group mean curve; (k): number of individual response clusters.

The same statistical analyses were carried for children's H scores and R^2_{lin} coefficients. Regarding H scores, there was no main significant effects of experimental group ($F(1,22) = 1.464$, $p = .239$), or of test moment ($F(1,22) = 3.107$, $p = .092$). There was, however, a significant interaction effect ($F(1,22)= 7.661$, $p = .011$). H scores in the first number-to-position evaluation were higher in the Control group (M = .62 , SD = .18) than in the Experimental group (M = .47, SD = .19). In the second evaluation (post-test), H scores decreased in the Control group (M = .52, SD = .20), but increased slightly in the Experimental group (M = .49, SD = .22). The interaction effect is clarified both by addressing the response clusters (Figure 4) and the single-trial scatterplots (Appendix B).

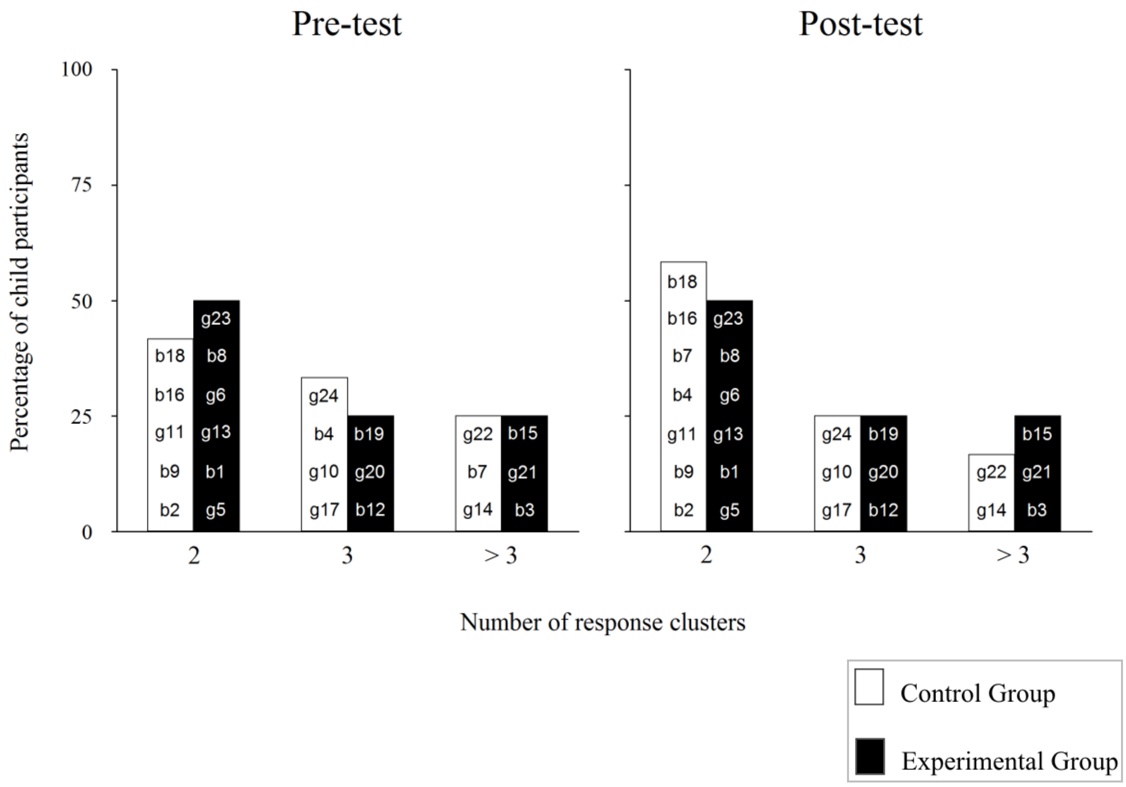


Figure 4. Percentage of participants whose single-trial distribution of responses along the extent of the bar conglomerated onto 2, 3, or more clusters.

As Figure 4 shows, ten participants in the Control Group maintained their cluster patterns in the Post-test Number-to-Position task. The exception, b4 and b7, responded at three or more response clusters when they were first tested in the Number-to-Position task (Pre-test) but in the Post-test responded bi-categorically which, in turn, led to a decrease in their H scores. As for the majority who retained the same number of clusters in Pre- and Post- test moments, four of them distributed their responses more evenly in the post-test (higher H scores), while the other six distributed their responses less evenly (lower H scores). However, the overall decrease in the Control group's post-test H scores was not statistically significant.

All participants in the Experimental Group maintained their cluster patterns (see black bars in Figure 4). Only three participants in the Experimental group presented lower H scores in the post-test, against the remaining nine whose H scores are higher in the post-test than in the pre-test. Inspection of the single trial scatterplots (Appendix B) does seem to suggest a broader and more evenly distribution of responses along the extent of the bar but, as the ANOVA analysis revealed, this difference between test moments was not statistically significant.

Lastly, the statistical test on the coefficients of determination yielded by the fitting of a simple linear function to each participant's mean response location (R^2_{lin}), revealed no significant effect of the experimental group ($F(1,22) = .003$, $p = .955$), of the assessment moment ($F(1,22) = 1.232$, $p = .279$) and neither an interaction between the two factors ($F(1,22) = 1.207$, $p = .284$).

4. Conclusions

The main question in the current study was whether categorization training would alter perceptual judgments measured by a number-to-position task. We speculated that, if we implement a training to increase perceptual sensitivity to the numerosities, subjects would be less likely to assign them to overlapping locations along the bar. Alternatively, the training effect could make the Number-to-position instructions clearer to preschoolers. For instance, the specific visual tagging of numerosities along the tested range could lead preschoolers to formulate rules such as "because numbers are different, they belong to different locations".

The categorical training protocol that we implemented did not seem to affect adults' Number-to-Position performance. Firstly, adult participants had already understood the instructions, for they responded continuously along the bar in the first

Number-to-position task. Secondly, the treatment did not affect how evenly (H scores) or linearly (R^2_{lin} coefficients) the responses were distributed along the bar.

Seventy five percent of the preschoolers responded bi- and tri-categorically when they were first tested in the Number-to-Position task, in both the Control and Experimental groups. The percentage of categorical participants in the Control group increased to 83.3% in the post-test, whereas all Experimental group participants maintained their previous number of response clusters. In the post-test, Normalized entropy (H) scores decreased in the Control group, but increased slightly in the Experimental group which suggested that the pre-training made participants distribute their responses more evenly among the clusters. However, the participants did not select more clusters. As such, we conclude that the perceptual training within the tested numerical range had a very modest effect, for although it made participants distribute their responses more evenly among the clusters, it did not made participants select more clusters than in the pre-test. In other words, the training did not seem to help children understand that they were to respond along the bar, when they were re-tested in the Number-to-Position procedure.

We cannot dispute that the training protocol could have been implemented differently, and that some other parameterizations in the variables known to affect other stimulus dimensions and/or categories could have yielded different outcomes. For example, variables such as amount of practice (Karni & Sagi, 1993; Qu, Song, & Ding, 2010), the duration of the interval between the categorical training and the assessment task (Özgen & Davies, 2002), and how feedback is introduced (Gibson & Gibson, 1955; Herzog & Fahle, 1998) have all been shown to affect perceptual learning. There is yet another dynamic feature of the training protocol that has been reported since the famous study of Ivan P. Pavlov and his student, Nataliia Shenger-Krestovnikova, with dogs trained to discriminate circles from ellipses (Pavlov, 1927). This finding, which was later systematically investigated and named “transfer of a discrimination along a continuum” by Lawrence (1952), is that “a difficult discrimination is more easily established if the subjects are first trained on an easy discrimination of the same type than if all the training is given directly on the difficult discrimination” (Lawrence, 1952; Sutherland & Mackintosh, 1971).

Although it does not fall in the scope of the current study, it would be interesting to observe if sensitivity to number or, in Gibson’s terminology, the “differentiation of the percept” (Gibson & Gibson, 1955) would change as a condition of the task applied to assess numerical discrimination. From categorical perception literature or even the broader human psychophysics field, we have reason to expect an effect of the type of task (e.g., Laming, 1984, p. 161; Angulo & Alonso, 2012).

Interestingly, in a parallel experiment we carried out to explore the effects of categorical training and mere exposure on nonverbal number bisection performance of adults, sensitivity in the [10,90] range was not affected by the pre-training (for detailed description, this study is reported in Appendix E of the current thesis). A possible reason was that the few/many anchor ratio studied, 1:9, is too easy a discrimination for adults (Halberda & Feigenson, 2008). We found that many individual psychometric functions were categorical, in the sense that instead of a gradual monotonic function, probability of “many” changed abruptly from 0 to 100%, in a step-like manner, and perhaps they were the result of “all-or-none” rules of responding such as «if sample is not ‘10’, then answer “many”». Nevertheless, we believe that, by testing with smaller ranges and ratios, it would be interesting to investigate the effect of categorical training on humans’ numerical discrimination, as it may be assessed with diverse nonverbal procedures, such as small/different judgment tasks, “choose the largest” type of comparisons, and bisection tasks.

CHAPTER VI : GENERAL DISCUSSION

General Conclusions

The purpose of our first number-to-position experiment, presented as Study 1, was to extend Dehaene and colleagues (2008)'s procedure to the study of preschoolers' number to space mapping. To our knowledge, up until now, prior preschoolers' Number-to-position (NTP) studies had tested solely symbolic stimulus conditions (spoken number words and/or written Arabic digits). To that end, preschoolers and adults were tested in NTP tasks with sets of dots in the [1,10] and [10,100] ranges, sequences of [1,10] tones and spoken Portuguese numerals [1,10]. However, differently from Dehaene and colleagues' study, to reduce possible biases due to the experimenter-participant interaction, our procedure was fully computerized.

The most important novelty of Study 1 was that we inspected individual data, in addition to group data. We found that different levels of analysis led to conflicting conclusions. On one hand, preschoolers' group average curves replicated the patterns found in the literature, namely, that response location increased monotonically with the numerical sample, in a linear-like pattern for symbolic stimuli (number words) and a logarithmic-like pattern for nonsymbolic stimuli. On the other hand, inspection of single trial distribution of responses revealed that preschoolers exhibited distinct response strategies. Most important, we identified preschoolers who restricted their responses to the two trained anchor positions (bi-categorical pattern), preschoolers who responded at the bar's endpoints and midpoint locations (tri-categorical), and participants who presented a broader distribution along the bar (continuous). The importance of the single-trial inspection was that it revealed that many preschoolers' performance violates the own assumption of the number-to-position procedure, which is that, regardless of the spacing between consecutive numbers, their spatial positioning must obey their ordinal relationship (Siegler & Opfer, 2003; Núñez, Cooperrider, & Wassman, 2012).

Since adults' individual mapping patterns are unquestionably continuous, how do the initial categorical patterns become as such? Which prior learnings might have enhanced preschoolers' performance? These questions determined the design of the three remaining experiments. In Studies 2, 3 and 4, we manipulated participants' recent pre-training history and tested its effects on preschoolers' propensity to respond along the spatial response continuum.

Improvement in NTP performance was mostly evaluated by classifying subjects according to the number of their response clusters along the bar (2, 3, and >3 response clusters), as well as by Normalized Entropy (H) scores. The latter is a measure typically employed in Information Theory (Shannon, 1948; Shannon & Weaver; 1949). Because the NTP literature is more concerned with average group data, entropy has not been used to describe individual NTP performance (though see, e.g., Young & Wasserman, 1997, 2001, 2007, where it was employed to describe pigeons' discrimination of stimulus variability). In a [10,90] NTP situation, considering the relative frequencies of responses distributed into nine bins, a perfectly even distribution would lead to the maximum H score, 1. However, the more skewed the distribution of responses along the bar (e.g., a bi-categorical pattern), the lower the H score. We found that the pre-trainings increased H scores, in comparison with Control groups, in which participants were solely tested in the NTP task. Yet, the extent of such improvement greatly depended on the pre-training manipulation.

The largest difference between the Experimental (with pre-training) and Control (no pre-training) groups occurred in Study 2. The pre-training in a Brightness-to-Position task experienced by the Experimental group led to significantly higher H scores and, in the Dots[10,90] range, more than thrice the number of ">3"cluster patterns than the Control group.

In the Figures-to-Position pre-training, underwent by participants in the Experimental group of Study 3, we isolated the effect of having responded in a spatial *continuum* prior to the NTP testing. Although the number of bi- and tri- categorical patterns was smaller and normalized entropy was higher in the Experimental group, the differences between groups were not significant.

The main difference between the pre-training implemented in Study 2, the Brightness-to-position pre-training, and the Figures-to-Position pre-training in Study 3, was that in the latter the samples were nor ordered along a sensory *continuum*. The higher transfer from the pre-training to the NTP testing must thus depend on the degree of similarity between the stimulus and response components of the two performance situations. Specifically, in both the Brightness-to-Position training and the Number-to-Position test sessions, the stimuli were ordered in a continuum of increasing magnitude, as well as the response continuum. In other words, contrary to the Figures-to-position task, the Brightness-to-Position and Number-to-Position tasks may be labeled as continuous repertoire tasks (Scheuerman, Wildemann, & Holland, 1978; Rosenberg, 1963; Wildemann & Holland, 1972).

However, because half the participants in Study 2 were tested in the Dots[1,9] range before being tested in the [10,90] range, we cannot rule out the influence of this variable. In other words, the testing with smaller, easily countable numerosities might have helped children's to understand the instruction that larger numerosities were to be "placed" along the bar as well. In hindsight, we should have tested for this order effect. At the very least, in Study 3 we should have added a third experimental group which would have been solely tested in NTP tasks, but on the Dots[1,9] condition before the Dots[10,90] one.

The experimental design in Study 4 differed from the previous in that it was a case of a pretest-posttest control-group research design, with two experimental groups. Specifically, each participant was tested twice in Dots[10,90] NTP tasks, but in-between these testing moments participants in the Experimental Group completed a perceptual training protocol. In the post-test NTP evaluation, H scores from the Control group decreased and H scores form the Experimental group increased. This was at most a very modest change, for although participants in the Experimental group distributed their responses more evenly among the clusters, in no group did the participants select more clusters than those presented in the pre-test.

To summarize, we found that preschoolers' use of the spatial continuum in a NTP task increased if they had previously experienced a pre-training of mapping non-numerical continuous stimuli (Study 2: brightness). Improvement was also found, although to a much lesser extent when participants had to discriminate numerosities within the range tested in the NTP task (Study 4: perceptual training). Specifically, there was a modest increase in how evenly responses were distributed (H scores). Finally, learning to select different locations along the bar as a function of non-ordered, arbitrary stimuli (Study 3: TV cartoon characters) did not help children respond along the bar.

Preschoolers' categorical patterns and the "mental number line" hypothesis

Young children seem to have the ability to map numbers continuously (Experimental groups), even though without the pre-training they may not exhibit such ability (Control groups). Older participants do not require the training to do so. But certainly such age differences are a question of how readily the ability is displayed and not to be interpret as if children "lack either the capacity for the task in question or the skills needed to show that capacity" (Gelman & Gallistel, 1978). What needs to be questioned is the nature of this ability. In fact, we see no advantage in treating it as an "automatic", "innate", "inherent" or "privileged" predisposition to assign numbers onto space. Rather, we conceive it as an acquired continuous repertoire (Holland & Skinner, 1961; Rosenberg, 1963).

Núñez, Doan, and Nikoulina (2011) have also presented a case against the "mental number line" account, with actual mapping experiments. The authors proposed that assigning numbers onto space is no different than assigning them to other non-spatial response continuums. In their study, American undergraduate students were tested in numerical estimation tasks, with half of the participants tasked with reporting numbers spatially and the other half nonspatially. The spatial reporting condition was a NTP task, where a line segment was depicted on screen, flanked by anchor sets of dots, similarly to Dehaene et al. (2008)'s NTP procedure. Also similarly to Dehaene and colleagues (2008)' study, numerosities could be presented symbolically (Numerals[1,10]) and nonsymbolically (Dots[1,10], Dots[10,100], Tones[1,10]). The three nonspatial reporting conditions involved squeezing a dynometer, striking a cowbell with a soft-tipped mallet, and producing vocalizations of different intensities. The authors plotted mean perceived response intensity as a function of numerosity. It was found that the patterns observed in the spatial reporting condition (NTP) were practically reproduced with nonspatial reporting conditions. The only difference was that the Dots[1,10] presentation yielded a linear-like mapping in the NTP report condition, but a logarithmic-like one in all the nonspatial report conditions. The authors attributed this difference to the extensive exposure to external number lines within the easy to count [1,10] range, which arguably leads to a comparatively higher level of precision of the spatial estimates. But the high similarity between the estimation curves obtained across different report conditions suggests that number representation is "neither monolithic nor fundamentally spatial" (Núñez, Doan, & Nikoulina, 2011; see also Cantlon, Cordes, Libertus, & Brannon, 2009). The authors concluded that although "space provides a myriad of metaphors for number that pervades the history of modern mathematics", it does not demand for the assignment of numbers onto space to be "a universal intuition rooted in brain evolution, emerging early in ontogeny independently of education and culture (Dehaene et al., 2008)".

Throughout each study's discussion of results, we have established the argument for how preschoolers' performance in the NTP task should not be taken as a direct readout of the psychological representation of numbers (e.g., a *mental number line*). However, we must still explain why most children who responded bi- or tri-categorically demonstrated sensitivity to number, as shown by proportion of right-endpoint responses increasing with numerosity.

As we have previously mentioned, the bi-categorical pattern is alike to responding in a numerical bisection procedure, as if subjects treated the bar's endpoints as two response *manipulanda*: the leftmost endpoint for "Few" and the rightmost endpoint for "Many" responses. The usual result in bisection experiments is that the proportion of "Many" responses increases monotonically with numerosity, in a sigmoid-like fashion. This psychometric curve is usually the basis for further analyses. A parameter of interest is the Bisection Point (BP; also called Point of Subjective Equality or Point of Subjective Indifference), the numerosity at which the subject responds equally to both *manipulanda* (Proportion("Many") = 0.5). In other words, the bisection point is the numerosity perceived as being equally distant from the two anchor values, in a subjective (i.e., psychological) scale. Numerosities smaller than the bisection point are perceived as more similar to the 'Few' anchor, whereas those above it are perceived as more similar to the 'Many' anchor value.

The Bisection Point may be useful to measure how certain experimental manipulations may affect (e.g., augment) the perception of the numerosities. Yet another aspect about the BP is that it can be taken as a measure of the format of the psychological scale of numerosity. A BP closer to the arithmetic mean (AM) of the anchor values (AM = (few + many)/2) is taken an indicator of a precise linear scale with constant spacing between subsequent numbers together with constant generalization or variability associated with each number (see, e.g., Figure 1 in Cantlon et al., 2009 for an illustration of the three forms of subjective numerical scales). On the other hand, a bisection point at the geometric mean (GM) of the anchors (GM = √(few × many)) would be predicted assuming either a logarithmic spacing with constant generalization around all numbers (as defended by, e.g., Dehaene & Changeux, 1993), or a linear spacing with generalization increasing proportionally with number (e.g., Meck & Church, 1983).

In light of this reasoning, the data analyses usually carried with Bisection data could be applied to the bi-categorical cases found in NTP tasks, which would allow us to gauge the format of the subjective scale for numerosities. However, there is reason to believe that even numerical and temporal discrimination behavior of animals and humans does not just reflect a "psychophysical" transformation of physical/objective numerical quantity. In fact, some authors have proposed the mediation of response rules (or response thresholds) or an association between representations of number and response classes (e.g., Wearden, 1992; Droit-Volet, Clément, & Fayol, 2003; Droit-Volet & Izaute, 2009; Tan, 2009, p. 312; Almeida, Arantes, & Machado, 2007).

Thus, the question is how behavioral estimates (i.e., responses on the bar) may reveal the transformation of physical numbers into a psychological representation (Shepard, 1981), as is illustrated in Figure 1.

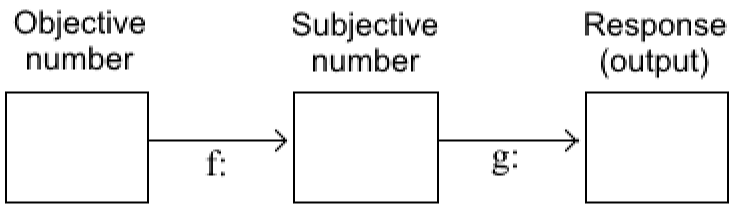


Figure 1. Schematic illustration of the psychophysical transformation of objective (physical) number into a psychological magnitude (f:) and a second transformation into an externally observable response (g:).

Physical (objective) numbers are perceived according to a psychophysical transformation, f:. This function may lead to a power (Stevens, 1957), a logarithmic with constant variability (Dehaene & Changeux, 1993) or a linear with proportional variabilty (Meck & Church, 1983) representation of numerosity. This subjective numerosity is then transformed (g:) into an externally observable response.

Applied to the Number-to-position situation, the authors who advocate the mental number line assume that f: transforms objective number into a spatial representation of numbers (*mental number line*), in which numbers are spacially ordered by increasing magnitude, in a logarithmic fashion. Crucially, the transformation between subjective number and external responses is an identity function ($y(x) = x$). This would explain the author's proposal that positions along the bar mirror the (unobservable) psychological representation of numbers.

However, the formulation as it is, cannot account for children's bi- and tri-categorical patterns of response, as they were observed in our NTP studies. The possibility of category-based numerical discriminatios is illustrated in Figure 2.

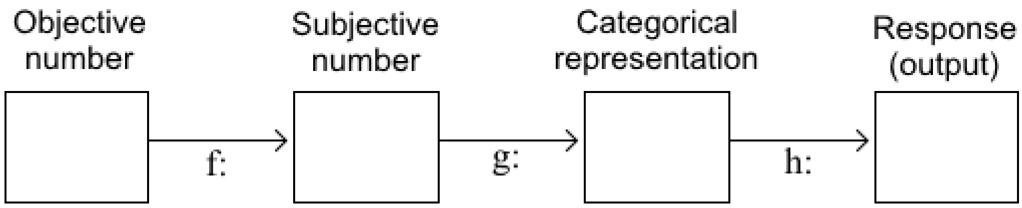


Figure 2. Schematic illustration of the psychophysical transformation of objective (physical) number into an externally observable response. In-between the subjective representation and the behavioral response, subjective numbers are transformed into categories or response classes (function g:).

To account for the SNARC effect, Santens and Gevers (2008) have proposed such category-mediated performance. In their model, numerical samples are represented as categorized, from the lesser to the larger number (Gevers, Verguts, Reynvoet, Caessens, & Fias, 2006). The authors propose that the intermediate representation bridges the pure numerical representation and the spatial response dimension (respond left/right, bottom/top, close/far) (illustrated as function g:, Figure 2). Also, it is the specific connections between the intermediate categorical representations and the responses available that are dependent on the procedural manipulations. Santens and Gevers (2008)' model could be adapted to the Number-to-Position tasks, assuming that initial representations of number are mapped into categorical discriminations of numerosity, in an ordinal scale of "few", "medium" and "many". Perhaps the number of categories could vary. Two categories could result in the bi-categorical mapping, of "Few" vs. "Many" responses; three categories in the "few", "medium" and "many" response pattern, and so on.

A favorable aspect of this possibility is that it is reunites NTP findings with an extensive literature on perceptual proportion judgments (Spence, 1990; Hollands & Dyre, 2000; Hollands, Tanaka, & Dyre, 2002). This endeavor has already started with Barth and Paladino (2011)'s study, the only that inspected children's individual scatterplots in symbolic (Arabic digits and spoken words) NTP tasks. The authors' proposed model of proportional-judgment not only accounts for individuals' NTP performance, it also replicates prior symbolic NTP studies' group average results. Furthermore, the model accounts for the reported log-to-linear developmental transition in estimation by assuming that the number of reference points (hallmarks) increases with age. For example, a younger child may use only the anchors as reference points, but older children may start taking into account a central reference point, and so forth.

Can NTP performance predict the future acquisition of Arithmetic?

An assumption pervading the Number-to-Position (NTP) literature, is that young children whose spatial mapping is more linear than logarithmic tend to learn arithmetics more easily whereas, in comparison, children with a more logarithmic-like mapping are at a disadvantage (Siegler & Ramani, 2009). This theoretical assumption has lead researchers to suggest that the Number-to-position task may prove a useful tool to screen for potential future disabilities in the acquisition of complex abilities such as Arithmetic (Siegler & Booth, 2004; Siegler & Ramani, 2008, 2009).

The attempt to implement a numerical discrimination task as a screening tool for future learning frailties is not unusual in the field of human numerical cognition and it certainly is not exclusive to the Number-to-position task. In fact, it seems that the idea to use the NTP procedure stems from the series of experiments that followed a famous study published in 2008 in Nature (Halberda, Mazzocco, & Feigenson, 2008). In this developmental study, children were evaluated in the time period between pre-school and first high school year. The main finding was that, among variables such as individual standardized intelligence scores, visual reasoning, spatial reasoning, and other tests' scores, the variable that better predicted mathematical academic success was the participant's sensitivity in a "more vs. less" comparison task with visual, nonsymbolic numerical arrays ("are there more blue or yellow circles?", http://www.panamath.org/testyourself.php). This result has led many authors to propose the implementation of trainings in simple numeral discriminations with young children (Mazzocco, Feigenson, & Halberda, 2011a, 2011b; Wilson, Dehaene, Pinel, Revkin, Cohen, & Cohen, 2006; Wilson, Revkin, Cohen, Cohen, & Dehaene, 2006 (http://www.thenumberrace.com/nr/home.php); Butterworth & Laurillard, 2010 (http://www.low-numeracy.ning.com)).

However, we believe there are many obstacles to the possibility of, given the data collected thus far, implementing the NTP procedure with this aim of evaluating and screening for possible future learning disabilities, as it has been the case with the "more vs. less" task. A first obstacle is that, contrary to "more vs. less" discrimination tasks, there are considerably fewer NTP studies.

More important, we have yet to ascertain how certain procedural features may differentially affect children and adults' performance. Take the level of data analysis, for example. We still don't know the extent to which results obtained with preschoolers' NTP tasks with Arabic digits may be an averaging artifact. Moreover, we still do not know how performance in NTP task relates to the sensitivity to number and space, as measured in other discrimination procedures. Questions about stimulus modality are also of extreme importance when one aims to theorize about the cultural component of a hypothetical number-space relationship. As for the crucial type of stimulus conditions - when numerosities are presented non symbolically - no study has yet tested how children from different ages perform in non-symbolic NTP tasks. Only resorting to immediate nonsymbolic stimuli, which do not require cultural learnings and thus may be presented to several animal species and human subgroups, might reveal a possible association between numerical stimuli and the distribution of responses along a spatial continuum.

By the time he/she is 5-years old, it is possible that the child has already been exposed to numerals ordered in a line and has utilized marked rulers with numbers from 1 to 10, oriented from left to right. An alternative approach to study a supposedly privileged number-space association while trying to decrease the contribution of such exposure, which can't be controlled by the experimenter, is to opt for different formats of spatial responses. One could test this contribution by studying, for example, how responses are distributed if the line is ordered from right to left. In another manipulation, one could present participants with a vertical response bar, in which numbers could be ordered from top to bottom or bottom to top. Finally, we need to contrast the performance of children from different countries and, more important, from different writing and reading orientation cultures. According to the mental number line hypothesizes, this variable should affect performance (Gobel, Shaki, & Fischer, 2011; Shaki, Fischer, & Petrusic, 2009; Zebian, 2005).

Future studies ought to ascertain what skills the child brings to the task, including her knowledge about the concepts used in the verbal instructions. For example, the apparatus may be presented to the participant as "*a path that goes from the smallest to the largest number. The smallest number belongs to the beginning and the largest to the end of the path. The larger the number, the more one must advance along the path*". What understanding do young children have of the relations implied in the instruction? Do they learn about them and measurement as well as about the variations in quantity as they experience successive trials (Gelman & Gallistel, 1978. pp.22-23.)? Do results differ when the task is presented as a hypothetical story, such as when the child is told that the response bar's positions signify the number of chocolates pieces required for a number of guests invited to a birthday party (Ebersbach, Luwel, Frick, Onghena, & Verschaffel, 2008)? What formulations of the instructions better encourage the child to respond along the extent of the bar? Which quantity expressions are more adequate for certain age groups (Gelman & Gallistel, 1978; Halberda, Taing, & Lidz, 2008)? Will different expressions lead to different mapping patterns? Given that young children's verbal repertoire to express quantification and relational rules is more limited than adults' (Cordes & Gelman, 2005), one must consider the potential of instructions to mislead the child and the context in which they are given (Dehaene, 1997, p.45). The importance of these variables is illustrated by the fact that they may account for the different studies' conclusions about children's success in Piaget (1952)'s number conservation tests (Mehler & Bever, 1967; McGarrigle & Donaldson, 1975; Gelman, 1969; Gelman & Gallistel, 1978).

Another aspect in need of attention is replicating those studies conducted with children in which trial events are highly dependent on the presence of and prompts from the experimenter. Prior NTP Procedures with young children entailed substantial intervention from the experimenter (Barth & Paladino, 2011; Berteletti et al., 2010). When presenting the task, the experimenter will sometimes exemplify to the child how to respond, as in observational learning or "modeling" situations (VandenBoss, 2006; Mazur, 2002, pp. 282-303; Meichenbaum, 1977; Welkowitz, & Calkins, 1984). The experimenter remains seated next to the child, and guides her throughout the trial events. He sometimes reads the Arabic digit aloud to the participant, poses the question of where to place it along the line, and provides verbal feedback following a response, such as "good job" and "thank you" (Siegler & Opfer, 2003; Siegler & Booth, 2004; Opfer & Siegler, 2007; Ebersbach et al., 2008; Berteletti, Lucangeli, Piazza, Dehaene,

& Zorzi, 2010; Barth & Paladino, 2011).In Dehaene and colleagues (2008)' study with the Munducuru, it was also the experimenter that registered the response with the mouse device, at the location pointed by the participant with his/her finger. It is not uncalled for to suspect that these interactions, with prompts and possible incidental cues from the experimenters' verbal and emotional responses, may influence the child's performance, especially with younger groups (e.g., Meichenbaum, 1977; Welkowitz & Calkins, 1984; Demchak, 1990; MacDuff, Krantz, & McClannahan, 2001). We cannot however determine up to which point the experimenter has influenced them.

Finally, although in all studies with children the sessions start with the participants being trained with the small-leftmost and large-rightmost mappings, we have yet to evaluate the contribution of the additional training about the middle number and position. The middle number-midpoint location training was implemented in some studies, but was never systematically varied so we wonder if this additional reference promoted a more linear-like mapping or, more probably, a tri-categorical response pattern. One can suppose that training this midpoint mapping, in addition to the anchors, might increase the number of individual tri-categorical patterns, a hypothesis also hinted at in the only Arabic digits NTP task which has depicted individual scatterplots (Barth & Paladino, 2011).

In addition to our ignorance of how parameterization and other procedural details affect performance, another type of problem pertains to data analysis. Prior NTP studies restricted their data analysis to the average group curves and, less frequently, to individual average curves. As a result, one cannot rule out the hypothesis that even in prior Arabic digits NTP studies the monotonically increasing response location curves constituted an averaging artifact. What does a preschooler, presented with 1-100 or 1-1000 Arabic digits number-to-position tasks, know about a numerical sample such as "87"? Can it be that if we address single-trial data in symbolic NTP tasks we will find that participants respond at the endpoint and middle point locations (i.e., categorically), as we found in our nonsymbolic studies? Theoretically, the hypothesis that numbers are spontaneously mapped onto space (de Hevia & Spelke, 2009, 2010) is challenged to account for categorical patterns. At the very least, the possibility that categorical patterns were "swept" due to averaging casts some reservation as to whether "the notion of automatic spatial coding of numbers" (Zorzi et al., 2006) should be evaluated with number-to-position procedures.

In conclusion, we believe that the major contribution of the studies in the current thesis was the construction of a case for the importance of corroborating whether group curves depict participants' true response patterns. The evidence we have collected, jointly with critical revisions of previous studies (e.g., Núñez, 2011), stress the need for a "debugging" of the NTP procedure.

REFERENCES

“mapping” . In Encyclopædia Britannica website. Available: Retrieved October 2, 2014, from http://www.britannica.com/EBchecked/topic/363594/mapping.

Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention and transfer of human motor skills. *Psychological Bulletin*, 101, 41-74.

Ahissar, M., Laiwand, R., & Hochstein, S. (2001). Attentional demands following perceptual skill training. *Psychological Science*, 12 (1), 56–62.

Allan, L. G. (2002). The location and interpretation of the bisection point. *Quarterly Journal of Experimental Psychology B*, 55, 43-60.

Allman, M. J., Pelphrey, K. A., & Meck, W. H. (2012). Developmental neuroscience of time and number: implications for autism and other neurodevelopmental disabilities. *Frontiers in Integrative Neuroscience*, 6(7), 1-24.

Almeida, A., Arantes, J., & Machado, A. (2007). Numerosity discrimination in preschool children. *Journal of the Experimental Analysis of Behavior, 3* (88), 339–354.

Alsop, B., & Honig, W. K. (1991). Sequential stimuli and relative numerosity discriminations in pigeons. *Journal of Experimental Psychology: Animal Behavior Proce*sses, 17, 386-395.

Angulo, R., & Alonso, G. (2012). Human perceptual learning: The effect of pre-exposure schedule depends on task demands. *Behavioural Processes*, 91 (3), 244–252.

Anobile, G., Cicchini, G. M. & Burr, D. C. (2012). Linear mapping of numbers onto space requires attention. *Cognition, 3* (122), 454-459.

Anscombe, F. J. (1973). Graphs in Statistical Analysis. *American Statistician*, 27 (1), 17–21.

Ashcraft, M. H., & Moore, A. M. (2012). Cognitive processes of numerical estimation in children. *Journal of Experimental Child Psychology*, *111*, 246–267.

Bächtold, D., Baumüller, M., & Brugger, P. (1998). Stimulus–response compatibility in representational space. *Neuropsychologia*, 36, 731–735.

Baer, D. M., Peterson, R. F., and Sherman, J. A. (1967). The development of imitation by reinforcing behavioral similarity to a model. Journal of the Experimental Analysis of Behavior, 10, 405-416.

Baid, J.C., & Noma, E. (1975). Psychophysical study of numbers: I. Generation of numerical responses. *Psychological Research*, 37, 281-297.

Baird, J.C., Lewis, C., & Romer, D. (1970). Relative frequencies of numerical responses in ratio estimation. *Perception and Psychophysics*, 8, 358-362.

Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science*, *14, 125-135.*

Barth, H., Slusser, E., Cohen, D., & Paladino, A. (2011). A sense of proportion: Commentary on Opfer, Siegler and Young. *Developmental Science*, *14*, 1205-1206.

Benoit, L., Lehalle, H., & Jouen, F. (2004). Do young children acquire number words through subitizing or counting? Cognitive Development, 19, 291–307.

Beran, M. J., Decker, S., Schwartz, A., and Schultz, N. (2011). Monkeys (*Macaca mulatta* and *Cebus apella*) and human adults and children(*Homo sapiens*) enumerate and compare subsets of moving stimuli based on numerosity. *Frontiers of Comparative Psychology,* 2, 61.

Beran, M. J., Johnson-Pynn, J. S., & Ready, C. (2008). Quantity representation in children and rhesus monkeys: linear versus logarithmic scales. *Journal of Experimental Child Psychology*, 100(3), 225-233.

Bermejo, V. (1996). Cardinality Development and Counting. Developmental Psychology, 32 (2), 263-268.

Bermejo, V., Morales, S., deOsuna, J. G. (2004). Supporting children's development of cardinality understanding. Learning and Instruction, 14, 381–398.

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical Estimation in Preschoolers. *Developmental Psychology*, *46* (2), 545–551.

Bertillon, J. (1881). De la vision et la memoire des nombres. *La Nature*, 408, 202-203.

Bertillon, J. (1882). De la vision des nombres. *La Nature*, 265-268.

Bertillon. J. (1880). De la vision des nombres. *La Nature*, 378, 196-198,

Bickel, W. K., & Etzel, B. C. (1985). The quantal nature of controlling stimulus-response relations as measured in tests of stimulus generalization. *Journal of the Experimental Analysis of Behavior*, 44, 245-270.

Blough, D. S. (1975). Steady state data and a quantitative model of operant generalization and discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, 1, 3-21.

Boisvert, M.J., Abroms, B.D.,& Roberts, W.A. (2003). Human nonverbal counting estimated by response production and verbal report. *Psychological Bulletin Review*, 10, 683–690.

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *41*, 189-201.

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. Child Development, 79, 1016-1031.

Borges, M. A., & Sawyers, B. K. (1974). Common verbal quantifiers: Usage and interpretation. Journal of Experimental Psychology, 102 (2), 335-338.

Brannon, E. M. (2006). The representation of numerical magnitude. *Current Opinion in Neurobiology*, *16* (2), 222–229.

Brannon, E. M., Suanda, S., & Libertus, K. (2007). Temporal discrimination increases in precision over development and parallels the development of numerosity discrimination. *Developmental Science*, *6*, 770-777.

Brannon, E.M. (2006). The representation of numerical magnitude. *Current Opinion in Neurobiology*, 16, 222–229.

Briars, D., & Siegler, R. S. (1984). A featural analysis of preschoolers' counting knowledge. *Developmental Psychology*, 20, 607–618.

Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology*, 20, 493–523.

Buckley, P. B., & Gilman, C. B. (1974). Comparison of digits and dot patterns. *Journal of Experimental Psychology*, 103, 1131-1136.

Bulf, H., Macchi-Cassia, V.M., & de Hevia, M.D. (2014). Are numbers, size and brightness equally efficient in orienting visual attention? Evidence from an eye-tracking study. *PLoS ONE*, 9, e99499.

Burns, B. & Shepp, B.E. (1988). Dimensional interactions and the structure of psychological space: the representation of hue, saturation, and brightness. *Perception and Psychophysics*, 43, 494-507

Butterworth, B., & Laurillard, D. (2010). Low numeracy and dyscalculia: identification and intervention. *The International Journal on Mathematics Education*, 42(6), 527-539.

Cantlon, J. F. , Cordes, S., Libertus, M. E., & Brannon, E. M. (2009). Comment on 'Log or Linear? Distinct intuitions of the number scale in Western and Indigene cultures'. *Science*, 323 (38).

Cantlon, J.F., & Brannon, E.M. (2007). How much does number matter to a monkey (Macaca mulatta)? *Journal of Experimental Psychology: Animal Behavior Processes*, 33, 32–41.

Capaldi, E.J., & Miller, D.J. (1988) Counting in rats: its functional significance and the independent processes that constitute it. *Journal of Experimental Psychology: Animal Behavior Processes*, 14, 3-17.

Chelonis, J.J., Cox, A.R., Karr, M.J., Prunty, P.K., Baldwin, R.L., & Paule, M.G. (2014). Comparison of delayed matching-to-sample performance in monkeys and children. *Behavioural Processes*,103, 261–268.

Chelonis, J.J., Daniels-Shaw, J.L., Blake, D.J., & Paule, M.G. (2000). Developmental aspects of delayed matching-to-sample performance in children. *Neurotoxicology and Teratology*, 22 (5), 683–694.

Clearfield, M. W., & Mix, K. S. (1999). Number versus Contour Length in infants' discrimination of small visual sets. *Psychological Science*, 10 (5), 408-411.

Clearfield, M. W., & Mix, K. S. (2001a). Amount versus number: Infants' use of area and contour length to discriminate small sets. *Journal of Cognition and Development*, 2, 243–260.

Cohen-Kadosh, R. & Walsh, V. (2009). Numerical representation in the parietal lobes: Abstract or not abstract? *Behavioral and Brain Sciences*, *32*, 313-328.

Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bulletin & Review*, 18, 331–338.

Cohen-Kadosh, R., Lammertyn, J. & Izard, V. (2008) Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Progress in Neurobiology*, 84(2), 132–147.

Cordes, S. & Gelman, R. (2005).The young numerical mind: When does it count? In J. Campbell (Ed.), *Handbook of Mathematical Cognition*. London: Psychology Press, (pp. 127-142).

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.

Crites, T. (1992). Skilled and less skilled estimators' strategies for estimating discrete quantities. The Elementary School Journal, 92, 601–619.

Davis, H. (1984). Discrimination of the number three by a raccoon (Procyon lotor). *Animal Learning & Behavior*, 12, 409-413.

Davis, H. (1993). Numerical competence in animals: life beyond Clever Hans. In: Boysen, S.T., & Capaldi, E.J. (eds). *The development of numerical competence: animal and human models*. Erlbaum, Hillsdale, pp 109–125.

Davis, H., & Memmott, J. (1982). Counting behavior in animals: a critical evaluation. *Psychological Bulletin*, 92, 547–571.

Davis, H., Pérusse, R. (1988). Numerical competence in animals: Definitional issues, current evidence, and a new research agenda. *Behavioral Brain Science*, 11, 561–615.

de Hevia, M. D., & Spelke, E. S. (2009). Spontaneous mapping of number and space in adults and young children. *Cognition*, 110, 198–207.

de Hevia, M. D., & Spelke, E. S. (2010). Number-space mapping in human infants. *Psychological Science*, 21, 653–660.

de Hevia, M.D., Girelli, C., & Cassia, V.M. (2012). Minds without language represent number through space: origins of the mental number line. *Frontiers in Psychology*, 3, 466.

Dehaene, S. & Brannon, E.M. (2010a). Space, time, and number: A Kantian research program. *Trends in Cognitive Sciences*.

Dehaene, S. & Brannon, E.M. (2011). *Space, Time, and Number in the Brain: searching for the foundations of mathematical thought*. Elsevier.

Dehaene, S. & Brannon, E.M., Editors (2010b). Special Issue on Space, Time, and Number, *Trends in Cognitive Sciences*.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*, 1-42.

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

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

Dehaene, S. (2003). The neural basis of Weber-Fechner's law: Neuronal recordings reveal a logarithmic scale for number. *Trends in Cognitive Science*, 7, 145-147.

Dehaene, S. (2009). Origins of mathematical intuitions. *Annals of the New York Academy of Sciences, 1156*, 232–259.

Dehaene, S., & Changeux, J.P. (1993). Developmental of elementary numerical abilities: a neuronal model. *Journal of Cognitive Neuroscience*, 5, 390–407.

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, 43, 1–29.

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

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neuroscience,* 21, 355-361.

Dehaene, S., Dupoux, E., & Mehler, 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., Izard, V., & Piazza, M. (2005). Control over non-numerical parameters in numerosity experiments. Unpublished manuscript (available on www.unicog.org).

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

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008b). Supporting online material for "Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures." *Science Online*. Retrieved from www.sciencemag.org/cgi/content/full/320/5880/1217/DC1

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, 20, 487-506.

Demchak, M. A. (1990). Response prompting and fading methods: a review. *American Journal of Mental Retardation*, 94 (6), 603-615.

Descoeudres, A. (1921). *Le dévelopment de l'enfant de deux à sept ans*. Paris: Delachaux et Niestlé C.A

Dowker, A. (2008). Individual differences in numerical abilities in preschoolers. Developmental Science, 11, 650–654.

Droit-Volet S, Tourret S, & Wearden J. (2004). Perception of the duration of auditory and visual stimuli in children and adults. *Quarterly Journal of Experimental Psychology* A, 57 (5), 797–818.

Droit-Volet, S. & Zélanti, P (2013). Development of time sensitivity: Duration ratios in time bisection. *The Quarterly Journal of Experimental Psychology*, 66 (4), 671-686.

Droit-Volet, S. (2010). Speeding up a master clock common to time, number and length? *Behavioural Processes*, 85, 126-134.

Droit-Volet, S., & Izaute, M. (2009). Improving time discrimination in children and adults in a temporal bisection task: The effects of feedback and no-forced choice on decision and memory processes. *The Quarterly Journal of Experimental Psychology*, 60 (6), 1173−1188.

Droit-Volet, S., & Wearden, J. H. (2001). Temporal bisection in children. Journal of Experimental Child Psychology, 80, 142-159.

Droit-Volet, S., Clément, A., & Fayol, M. (2003). Time and number discrimination in a bisection task with a sequence of stimuli: A developmental approach. *Journal of Experimental Child Psychology*, *84*(1), 63-76.

Droit-Volet, S., Clément, A., & Fayol, M. (2008). Time, number and length: similarities and differences in discrimination in adults and children. *Quarterly Journal of Experimental Psychology, 61*(12), 1827-1846.

Ebersbach M., Luwel K., Frick A., Onghena P., Verschaffel L. (2008). The relationship between the shape of the mental number line and familiarity with numbers in 5- to 9-year-old children: evidence for a segmented linear model. *Journal of Experimental Child Psychology*, *99*, 1–17.

Emmerton, J., & Renner, J.C. (2006). Scalar effects in the visual discrimination of numerosity by pigeons. *Learning & Behaviour*, 34, 176–192.

Emmerton, J., & Renner, J.C. (2009). Local rather than global processing of visual arrays in numerosity discrimination by pigeons (Columba livia). *Animal Cognition*, 12, 511-526.

Eriksen, C. W., & Hake, H. W. (1955a). Absolute judgments as a function of stimulus range and the number of stimulus and response categories. *Journal of Experimental Psychology*, 49, 323-332.

Eriksen, C. W., & Hake, H. W. (1955b). Multidimensional stimulus differences and accuracy of discrimination. *Journal of Experimental Psychology*, 50, 153-160.

Estes, W.K., 1956. The problem of inference from curves based on group data. *Psychological Bulletin*, *53*, 134–140.

Fechner, G. T. (1889). *Elemente der Psychophysik* (2 Volumes) (2nd ed.). Leipzig: Breitkopf & Härtel.

Fernandes, D. M., & Church, R. M. (1982). Discrimination of the number of sequential events by rats. *Learning & behavior*, *10*(2), 171-176.

Fetterman, J.G. (1993). Numerosity discrimination: Both time and number matter. *Journal of Experimental Psychology: Animal Behavior Processes*, 19, 149-164.

Fischer, M. H. (2006). The future for SNARC could be stark. *Cortex*, 42, 1066–1068.

Fischer, M. H., Shaki, S., & Cruise, A. (2009). It takes only one word to quash the SNARC. *Experimental Psychology*, 56 (5), 361–366.

Fischer, M.H. (2001). Number processing induces spatial performance biases. *Neurology* 57, 822–826.

Foster, D. H. (1978). Visual comparison of random-dot patterns: Evidence concerning a fixed association between features and feature relations. *Quarterly Journal of Experimental Psychology, 30,* 637 -654.

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

Frye, D., Braisby, N., Lowe, J., Maroudas, C., & Nicholls, J. (1989). Young children's understanding of counting and cardinality. *Child Development*, 60(5), 1158-1171.

Fuson, K. C. (1988). *Children's counting and concepts of number*. New York: Springer-Verlag.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, 44, 43-74.

Gallistel, C.R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. Trends in cognitive sciences, 4(2), 59-65.

Galton, F. (1880a). Visualised Numerals. *Nature*, 252-256.

Galton, F. (1880b). Statistics of Mental Imagery. *Mind*, 5, 301-318.

Galton, F. (1881). Visualised numerals. *Journal of the Anthropological Institute*, 10, 85–102.

Galton, F. (1883). *Inquiries into human faculty and its development*, London: Macmillan.

Gandini, D., Lemaire, P., & Dufau, S. (2008). Older and young adults’ strategies in approximative quantification. *Acta Psychologica*, 129(1), 175–189.

Gebuis, T., & Gevers, W. (2011). Numerosities and space; indeed a cognitive illusion! A reply to de Hevia and Spelke (2009). *Cognition*, 121, 248-252.

Gelman, R. (1969). Conservation acquisition: A problem of learning to attend to relevant attributes. *Journal of Experimental Child Psychology*, 7, 167-187.

Gelman, R. (1972). The nature and development of early number concepts. In H. W. Reese (Ed.), *Advances in Child Development*, 3, New York: Academic Press.

Gelman, R. (1977). How young children reason about small numbers. In N.J. Castellan, D.B. Pisoni, & G.R. Potts (eds.). *Cognitive Theory*, 2, Hillsdale, New Jersey: Erlbaum.

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.

Gelman, R., & Tucker, M. F. (1975). Further investigations of the young child's conception of number. *Child Development*, 46, 167-175.

Gevers, W., Lammertyn, J., (2005). The hunt for SNARC. *Psychological Science*, 47, 10-21.

Gevers, W., Lammertyn, J., Notebaert, W., Verguts, T., & Fias, W. (2006). Automatic response activation of implicit spatial information: Evidence from the SNARC effect. *Acta Psychologica*, 122, 221–233.

Gevers, W., Verguts, T., Reynvoet, B., Caessens, B., & Fias, W. (2006). Numbers and space: A computational model of the SNARC effect. *Journal of Experimental Psychology. Human Perception and Performance*, 32, 32–44.

Ghirlanda, S. & Enquist, M. (2003). A century of generalization. *Animal Behaviour*, 66 (1), 15–36.

Gibbon, J. (1981). On the form and location of the bisection function for time. *Journal of Mathematical Psychology*, 24, 58-87.

Gibson, E.J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.

Gibson, E.J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.

Gibson, E.J. (1991). *An odyssey in learning and perception*. Cambridge, MA: MIT Press.

Gibson, J.J., & Gibson, E.J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review*, 62, 32-41.

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

Ginsburg, N. (1978). Perceived numerosity, item arrangement, and expectancy. *American Journal of Psychology, 91,* 267 -273.

Ginsburg, N. (1980). The regular-random numerosity illusion: Rectangular patterns. *Journal of General Psychology, 103, 211 - 216.*

Ginsburg, N., & Goldstein, S. R. (1987). Measurement of visual cluster. *American Journal of Psychology, 100,* 193 -203.

Göbel, S.M., Shaki, S., & Fischer, M.H. (2011). The Cultural Number Line: A Review of Cultural and Linguistic Influences on the Development of Number Processing. *Journal of Cross-Cultural Psychology*, 42(4), 543 -565.

Goksun, T., Woods, A. J., Chatterjee, A., Zelonis, S., Glass, L., & Smith, S. E. (2013). Elementary school children's attentional biases in physical and numerical space. The European Journal of Developmental Psychology, 10 (4), 433-448.

Goldstone RL (1998). Perceptual Learning. *Annual Review of Psychology*, 49, 585–612.

Goldstone RL, Lippa Y, Shiffrin RM. (2001). Altering object representations through category learning. *Cognition*, 78, 27–43.

Goldstone, R. L., & Hendrickson, A. T. (2009). Categorical Perception. Wiley Interdisciplinary Reviews: Cognitive Science, 1, 69-78.

Goldstone, R.L. (1994b). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, 123, 178-200.

Goldstone, R.L., & Hendrickson, A. (2010). *Categorical perception*. Wiley Interdisciplinary Reviews: Cognitive Science, 1(1), 69–78.

Gómez-Laplaza, L. M., & Gerlai, R. (2010). Can angelfish (Pterophyllum scalare) count? Discrimination between different shoal sizes follows Weber ' s law. *Animal Cognition*, 14 (1), 1-9.

Gray, S. A., & Reeve, R. A. (2014). Preschoolers' Dot Enumeration Abilities Are Markers of Their Arithmetic Competence. PLoS ONE, 9(4): e94428

Guttman, N., & Kalish, H. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology*, 51, 79-88.

Hake, H. W., & Garner, W. R. (1951). The effect of presenting various numbers of discrete steps on scale reading accuracy. *Journal of Experimental Psychology*, 42, 358-366.

Halberda, J., & Feigenson, L. (2008). Developmental Change in the Acuity of the "Number Sense": The Approximate Number System in 3- , 4- , 5- , and 6-Year-Olds and Adults. *Developmental Psychology*, 44 (5), 1457-1465.

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665-668.

Halberda, J., Taing, L., & Lidz, J. (2008). The Development of "Most" Comprehension and Its Potential Dependence on Counting Ability in Preschoolers. Language Learning and Development, 4(2), 99–121.

Hall, G. (1991). Perceptual and associative learning. Oxford, England: Clarendon Press.

Hall, G., (2001). Perceptual learning: Association and differentiation. In: Mowrer, R.R., Klein, S.B. (Eds.), *Handbook of contemporary learning theories*. Erlbaum, Mahwah, NJ, pp. 367–407.

Haun, D.B.M., Jordan, F.M., Vallortigara, G., & Clayton, N.S. (2010). Origins of spatial, temporal and numerical cognition: insights from comparative psychology. *Trends in Cognitive Science*, 14(12), 552–560.

Hauser, M. D., Dehaene, S., Dehaene-lambertz, G., & Patalano, A. L. (2002). Spontaneous number discrimination of multi-format auditory stimuli in cotton-top tamarins (Saguinus oedipus). *Cognition*, 86, B23-B32.

Herzog, M.H. & Fahle, M. (1998). Modeling perceptual learning: Difficulties and how they can be overcome. *Biological Cybernetics*, 78(2), 107-117.

Hinrichs, J.V., Yurko, D.S., & Hu, J.M. (1981). Two-digit number comparison: Use of place information. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 890-901.

Holland, J., & Skinner, B.F. (1961). *The Analysis of Behavior*. New York: McGraw-Hill.

Hollands, J.G., & Dyre, B.P. (2000). Bias in proportion judgments: the cyclical power model. *Psychological Review, 107*, 500–524.

Hollands, J.G., Tanaka, T., & Dyre, B. (2002). Understanding bias in proportion production. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 563–574.

Holloway, I.D. & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: the distance effect and children's mathematical competence. *Journal of Experimental Child Psychology*, 103 (1), 17-29.

Honig, W. K., & Stewart, K. E. (1989). Discrimination of relative numerosity in pigeons. *Animal Learning & Behavior*, 17, 134-146.

Honig, W. K., & Urcuioli, P. J. (1981). The legacy of Guttman and Kalish (1956): Twenty-five years of stimulus generalization research. *Journal of the Experimental Analysis of Behavior*, 36, 405-445.

Hourcade, J. P. , Bederson, B. B., Druin, A., & Guimbretiere, F. (2004). Differences in pointing task performance between preschool children and adults using mice. ACM *Transactions on Computer-Human Interaction*,11 (4), 357–386.

Howe, E., & Jung, K. (1987). Judgement of numerosity: Effects of symmetry and goodness in dot pattern arrays. *Acta Psychologica, 64,* 3-11.

Hubbard, E. M., Piazza, M., Pinel, P. & Dehaene, S. (2005) Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, *6* (*6*), 435–448.

Huntley-Fenner, G. (2001). Children's understanding of number is similar to adults' and rats': numerical estimation by 5‒7-year-olds. Cognition, 78, B27-B40.

Huttenlocher, J., Jordan, N. C., & Levine, S. C. (1994). A mental model for early arithmetic. Journal of Experimental Psychology: General, 123, 284–96.

Irie-Sugimoto, N., Kobayashi, T., Sato, T., & Hasegawa, T. (2009). Relative quantity judgment by Asian elephants (Elephas maximus). *Animal Cognition*, 12, 193-199.

Ito, Y., & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Memory & Cognition*, 32, 662–673.

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

James, W. (1890) *The principles of psychology*, volume I. New York: Dover Publications Inc., p 509.

Jenkins, S. E., & Cole, B. L. (1982). The effect of the density of background elements on the conspicuity of objects. *Vision Research, 22*, 1241 – 1252.

Jordan , K. E., Brannon, E. M., Logothetis, N. K., Ghazanfar, A. A. (2005). Monkeys match the number of voices they hear to the number of faces they see. *Current Biology*, 15, 1-5.

Jordan, K. E., & Brannon, E. M. (2006). A common representational system governed by Weber's law: Nonverbal numerical similarity judgments in 6-year olds and rhesus macaques. *Journal of Experimental Child Psychology, 95*, 215–229.

Jordan, K. E., Maclean, E. L., & Brannon, E. M. (2008). Monkeys match and tally quantities across senses. *Psychology*, 108, 617-625.

Jordan, K., & Brannon, E. M. (2006a). The multisensory representation of number in infancy. *Proceedings of the National Academy of Sciences*, 103(9), 3486-3489.

Jordan, K., & Brannon, E.M. (2006c). A common representational system governed by Weber's Law: Nonverbal numerical similarity judgments in six-year-old children and rhesus macaques. *Journal of Experimental Child Psychology*, 95, 215 229.

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

Karni, A. & Sagi, D. (1993). The time course of learning a visual skill. *Nature*, 365(6443), 250-252.

Klahr, D., & Walace, J. G. (1973). The role- of quantification operators in the development of conservation of quantity. *Cognitive Psychology*; .4, 301-327.

Konorski, J. and Miller, S. (1937). On two types of conditioned reflex. *Journal of General Psychology*, 16, 264-272.

Kopec, C.D., Brody, S.D. (2010). Human performance on the temporal bisection task. *Brain and Cognition*, 74, 262-272.

Krebs, J. R. (1978). Optimal foraging: Decision rules for predators. In J. R. Krebs & N. B. Davies (Eds.) *Behavioural ecology: An evolutionary approach*. Oxford, UK: Blackwell.

Krueger, L. E. (1984). Perceived numerosity: A comparison of magnitude production, magnitude estimation, and discrimination judgments. Perception and Psychophysics, 35(6), 536–542.

Laming, D. (1984). The relativity of 'absolute' judgements. *British Journal of Mathematical and Statistical Psychology*, 37, 152-183.

Landau, J.S. (1968). Postdiscrimination generalization in human subjects of two diferente ages. *Journal of Experimental Psychology*, 76, 656-663.

Landau, J.S. (1969). Size generalization in children as a function of test method and age. *Psychonomic Science*, 16, 57-58.

Landerl, K., Fussenegger, B., Moll, Kristina, & Willburger, E. (2009). Dyslexia and dyscalculia: Two learning disorders with different cognitive profiles. *Journal of Experimental Child Psychology*, 103, 309–324.

Laski, E. V., & Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development*, 76, 1723-1743.

Lawrence, D. H. (1952). The transfer of a discrimination along a continuum. *Journal of Comparative and Physiological Psychology*, 45, 511-516.

Le Corre, M. & Carey, S. (2008) One, two, there, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, *105*, 395-438.

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, 105, 395–438.

Lee, M. D., & Sarnecka, B. W. (2010). A model of knower-level behavior in number concept development. *Cognitive Science*, 34, 51–67.

Lipton, J. S., & Spelke, E. S. (2005). Preschool children's mapping of number words to nonsymbolic numerosities. *Child Development*, *76* (5), 978–988.

Longo, M. R., & Lourenço, S. F. (2007). Spatial attention and the mental number line: Evidence for characteristic biases and compression. *Neuropsychologia*, 45, 1400-1407.

Lourenco S. F., Longo M. R. (2010). General magnitude representation in human infants. *Psychological Science*. 21, 873–881.

Lu, C. & Frye, D. (1992). Mastering the Machine: A Comparison of the Mouse and Touch Screen for Children's Use of Computers. ICCAL, the International Conference on Computers and Learning (417-427). Wolfville, Canada: SpringerVerlag.

Lydersen, T., & Perkins, D. (1974). Effects of response produced stimuli upon conditional discrimination performance. *Journal of the Experimental Analysis of Behavior*, 21, 307-314.

MacDuff, G. S., Krantz, P. J., & McClannahan, L. E. (2001). Prompts and prompt-fading strategies for people with autism. In C. Maurice, G. Green, & R. M. Foxx (Eds.), *Making a difference: Behavioral intervention for autism* (pp. 37-50). Austin, TX: Pro-ed.

Machado, A., & Rodrigues, P. (2007). The differentiation of response numerosities in the pigeon. *Journal of the Experimental Analysis of Behavior*, *88*, 153 – 178.

Mazur, J. E. (2002). *Learning and behavior* (5th edition). Upper Saddle River, NJ: Pearson Education.

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011a). Impaired acuity of the approximate number system underlies mathematical learning disability (Dyscalculia). *Child Development*, 82(4), 1224–1237.

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011b). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PloS One*, 6(9), e23749.

McGarrigle, J., & Donaldson, M. (1975). Conservation Accidents. *Cognition*, 3(4), 341–350.

Mechner, F. (1958). Probability relations within response sequences under ratio reinforcement. *Journal of the Experimental Analysis of Behavior*, 1, 109-121.

Meck, W H, & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of experimental psychology. Animal behavior processes*, *9*(3), 320-34.

Meck, W. H., Church, R. M., & Gibbon, J. (1985). Temporal integration in duration and number discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, 11(4), 591–597.

Mehler, J., & Bever, T. G. (1967). Cognitive capacity of very young children. *Science*, 158, 141-142.

Meichenbaum, D. (1977) . *Cognitive-behavior modification: An integrative approach*. New York: Plenum Press.

Merten, K., & Nieder, A. (2009). Compressed scaling of abstract numerosity representations in adult humans and monkeys. *Journal of Cognitive Neuroscience*, 21 (2), 333-346.

Migler, B. (1964). Effects of averaging data during stimulus generalization. *Journal of the Experimental Analysis of Behavior*, 7, 303-30.

Minturn, A. L., & Reese, T. W. (1951). The effect of differential reinforcement on the discrimination of visual number. *The Journal of Psychology*, 31, 201–231.

Mix, K.S., Levine, S.C., & Huttenlocher, J. (1997). Numerical abstraction in infants: another look. *Developmental Psychology*, 33 (3), 423‑428.

Moxey, L. M., & Sanford, A. J. (2000). Communicating quantities: A review of psycholinguistic evidence of how expressions determine perspectives. Applied Cognitive Psychology, 14, 237-255.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature*, 215, 1519-1520.

Muldoon, K., Simms, V., Towse, J., Burns, V. & Yue, G. (2011). Cross-cultural comparisons of 5-year-olds' estimating and mathematical ability. *Journal of Cross Cultural Psychology*, 42, 669-681.

Müller, D., & Schwarz, W. (2007). Is there an internal association of numbers to hands? The task set influences the nature of the SNARC effect. *Memory & Cognition*, 35, 1151–1161.

Negen, J., & Sarnecka, B. W. (2010). Analogue Magnitudes and Knower-Levels: Re-Visiting the Variability Argument. In S. Ohlsson & R. Catrambone (eds.), Proceedings of the 32nd Annual Conference of the Cognitive Science Society. Paper presented at Cognitive Science Society, (1252 -1257). Red Hook, NY: Curran Associate.

Nieder, A., & Miller, E.K. (2003). Coding of cognitive magnitude.Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron, 37*, 149–157.

Nieder, A., & Miller, E.K. (2004). A parieto-frontal network for visual numerical information in the monkey. *Proc. Natl. Acad. Sci. USA, 101*, 7457–7462.

Nieder, A., Diester, I., & Tudusciuc, O. (2006). Temporal and spatial enumeration processes in the primate parietal cortex. *Science*, *313*, 1431-1435.

Notebaert, W., Gevers, W., Verguts, T., & Fias, W. (2006). Shared spatial representations for numbers and space: The reversal of the SNARC and the Simon effects. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 1197–1207.

Núñez, R. (2011). No innate number line in the human brain. *Journal of Cross-Cultural Psychology*, *42* (4), 651-668.

Núñez, R., Cooperrider, K., Wassmann, J. (2012). Number Concepts Without Number Lines in an Indigenous Group of Papua New Guinea. *PLoS ONE*, *7* (4): e35662.

Núñez, R., Doan, D, Nikoulina, A. (2011). Squeezing, Striking, and Vocalizing: Is Number Representation Fundamentally Spatial? *Cognition*, *120*, 225-235.

Opfer, J. E., & Furlong, E. E. (2011). How numbers bias preschoolers' spatial search. *Journal of Cross-Cultural Psychology*, 42 (4), 543 –565.

Opfer, J. E., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology*, 55, 169-195.

Opfer, J. E., & Thompson, C. A. (2006). Even early representations of numerical magnitude are spatially organized: Evidence for a directional magnitude bias in pre-reading preschoolers. In R. Sun & N. Miyake (Eds.), Proceedings of the 28th annual conference of the Cognitive Science Society (pp. 639–644). Vancouver, British Columbia: Cognitive Science Society.

Opfer, J. E., Siegler, R. S., & Young, C. J. (2011). The powers of noise-fitting: Reply to Barth and Paladino. *Developmental Science, 14*, 1194-1204.

Opfer, J. E., Thompson, C. A., & Furlong, E. E. (2010). Early development of spatial-numeric associations: Evidence from spatial and quantitative performance of preschoolers. *Developmental Science*, 13(5), 761–771.

Opfer, J., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology*, *55*, 169 – 195.

Ozgen, E., & Davies, I.R.L. (2002). Acquisition of categorical color perception: a perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General*, 131, 477–493.

Pavlov, I.P. (1927). *Conditioned reflexes*. Humphrey Milford: Oxford University Press.

Pepperberg, I. M. (1994). Numerical competence in an African Gray parrot (Psittacus erithacus). *Journal of Comparative Psychology*, 108, 36-44.

Perkins, D. N., & Salomon, G. (1992). Transfer of learning. In International Encyclopedia of Education (2nd ed.). Oxford: Pergamon.

Pfungst, O. (1911). *Clever Hans (The horse of Mr. Von Osten): A contribution to experimental animal and human psychology* (C. L. Rahn, Trans.). Retrieved from http://www.gutenberg.org/ebooks/33936.

Piaget, J. (1952). *The child's conception of number*. London: Routledge & Kegan Paul.

Piazza, M., Izard, V., Pinel, P., Le Bihan, D. & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, 44, 547–555.

Piazza, M., Mechelli, A., Price, C.J., & Butterworth, B. (2006). Exact and approximate judgements of visual and auditory numerosity: an fMRI study. *Brain Research*, *1106*, 177-188.

Pollack, I. (1952). The information of elementary auditory displays. *Journal of the Acoustic Society of America*, 24, 745-749.

Preacher, K. J. (2001, April). Calculation for the chi-square test: An interactive calculation tool for chi-square tests of goodness of fit and independence [Computer software]. Available from http://quantpsy.org.

Qu, Z. Song, Y & Ding, Y. (2010). ERP evidence for distinct mechanisms of fast and slow visual perceptual learning. *Neuropsychologia*, 48, 1869-1874.

Rattat, A. C., & Droit-Volet, S. (2012). What is the best and the easiest method of preventing counting in different temporal tasks? *Behavior Research Methods, 44*, 67–80.

Reeve, R., Reynolds, F., Humberstone, J., & Butterworth, B. (2012). Stability and Change in Markers of Core Numerical Competencies. *Journal of Experimental Psychology: General*. Advance online publication.

Rilling, M., & McDiarmid, C. (1965). Signal detection in fixed ratio schedules. *Science, 148,* 526-527.

Roberson D, Davidoff J, Davies IRL (2005). Color categories: evidence for the cultural relativity hypothesis. *Cognitive Psychology*, 50, 378–411.

Roberts, S. & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, 107 (2), 358-367.

Roberts, W. A ., & Mitchell, S. (1994). Can a pigeon simultaneously process temporal and numerical information? *Journal of Experimental Psychology: Animal Behavior Processes,* 20, 66-78.

Roberts, W. A ., Coughlin, R., & Roberts, S. (2000). Pigeons flexibly time or count on cue. *Psychological Science*, 11, 218-222.

Roberts, W. A. (1995). Simultaneous Numerical and Temporal Processing in the Pigeon. *Current Directions in Psychological Science, 4*(2), 47-51.

Roberts, W. A. (1998). *Principles of animal cognition*. Boston: McGraw-Hill.

Roberts, W. A. (2005). How do pigeons represent numbers ? Studies of number scale bisection. *Behavioural Processes*, *69*, 33-43.

Roberts, W. A. (2006). Evidence that pigeons represent both time and number on a logarithmic scale. *Behavioural Processes*, *72*, 207-214.

Roberts, W. A., Macuda, T., & Brodbeck, D. R. (1995). Memory for number of light flashes in the pigeon. *Animal Learning & Behavior*, 23, 182-188.

Roitman, J. D., Brannon, E. M., Andrews, J. R., & Platt, M. L. (2007). Nonverbal representation of time and number in adults. *Acta Psychologica*, *124*, 296-318.

Rosenberg, S. (1963). Behavior in a continuous-response task with noncontingent reinforcement. *Journal of Experimental Psychology,* 66 (2), 168-176.

Rousselle, L., & Noël, M. (2008). The development of automatic numerosity processing in preschoolers: evidence for numerosity-perceptual interference. *Developmental Psychology*, 44 (2), 544-560.

Rousselle, L., Palmers, E., & Noël, M-P. (2004). Magnitude comparison in preschoolers: What counts? Influence of perceptual variables. *Journal of Experimental Child Psychology*, 87, 57-84.

Routh, D. A. (1994). On Representations of Quantifiers. *Journal of Semantics*, 11, 199-214.

Santens, S., & Gevers, W. (2008). The SNARC effect does not imply a mental number line. *Cognition*, 108, 263–270.

Sasanguie D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number-space mappings: what underlies mathematics achievement? *Journal of Experimental Child Psychology*, 114, 418–443.

Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *British Journal of Developmental Psychology*, 30, 344–357.
Scheuerman, K. V., Wildemann, D. G., & Holland, J. G. (1978). A clarification of continuous repertoire development. *Journal of the Experimental Analysis of Behavior,* 30, 197-203.
Schleifer, P., & Landerl, K. (2011). Subitizing and counting in typical and atypical development. *Developmental Science*, 14 (2), 280-291.
Schmidt, R. (1991). Motor Learning & Performance: From Principles to Practice. Champaign: Human Kinetics Books.
Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, 47, 609-612.
Schwarz, W., & Keus, I. (2004). Moving the eyes along the mental number line: Comparing SNARC effects with manual and saccadic responses. *Perception and Psychophysics*, 66, 651664.
Seligman, M. E. P., & Meyer, B. (1970). Chronic fear and ulceration in rats as a function of unpredictability of safety. *Journal of Comparative and Physiological Psychology*, 73, 202-207.
Seron, X., & Pesenti, M., (2001). The number sense theory needs more empirical evidence. *Mind and Language*, 16(1), 76-88.
Seron, X., Pesenti, M., Nöel, M.P., Deloche, G., & Cornet, J.-A. (1992). Images of numbers, or "when 98 is upper left and 6 sky blue". *Cognition*, 44, 159-196.
Shaki S., Fischer M. H., & Petrusic W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychonomic Bulletin & Review*, 16, 328–331.
Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423; 623-656.
Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.
Shepard, R. N. (1958b). Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology*, 55, 509-523.
Shepard, R. N. (1981). Psychological relations and psychophysical scales: On the status of "direct" psychophysical measurement. *Journal of Mathematical Psychology*, *24*, 21-57.
Shettleworth, S. J. (2010). *Cognition, evolution, and behavior* (2nd Ed.). New York: Oxford University Press.
Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, *75*, 428 - 444.
Siegler, R. S., & Booth, J. L. (2005). Development of numerical estimation: A review. In J. I. D. Campbell (Ed.), *The handbook of mathematical cognition* (pp. 197−212). New York: Psychology Press.
Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation. *Psychological Science*, *14*, 237-243.

Siegler, R. S., & Robinson, M. (1982). The development of numerical understandings. *Advances in Child Development and Behavior*, 16, 241 - 312.

Siegler, R. S., Thompson, C. A., & Opfer, J. E. (2009). The logarithmic-to-linear shift: One learning sequence, many tasks, many time scales. *Mind, Brain, and Education*, 3, 143–150.

Siegler, R.S., & Ramani, G.B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental Science*, 11, 655–661.

Siegler, R.S., & Ramani, G.B. (2009). Playing linear number board games – but not circular ones – improves low-income preschoolers ' numerical understanding. *Journal of Educational Psychology*, 101(3), 545-560.

Simon, O., Mangin, J.F., Cohen, L., Le Bihan, D., & Dehaene, S. (2002). Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe. *Neuron*, 33, 475–487.

Simon, T. J., & Vaishnavi, S. (1996). Subitizing and counting depend on different attentional mechanisms: Evidence from visual enumeration in afterimages. *Perception and Psychophysics*, 58, 915–926.

Soltész, F., Szucs, D., & Szucs, L. (2010). Relationships between magnitude representation, counting and memory in 4- to 7-year-old children: A developmental study. *Behavioural and Brain Functions*, 6 (13), 1 - 14.

Spalding, J.M.K., & Zangwill, O.L. (1950). Disturbance of number-form in a case of brain injury. *Journal of Neurology, Neurosurgery and Psychiatry*, 13, 24-29.

Speelman, C. P., & McGann, M. (2013). How mean is the mean? *Frontiers in Psychology*, 4, 451.

Spence, I. (1990). Visual Psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 683–692.

Starkey, P., & Cooper, R. G. (1995). The development of subitizing in young children. *British Journal of Developmental Psychology*, 13 (4), 399–420.

Stevens, J. R., Wood, J. N., & Hauser, M. D. (2007). When quantity trumps number: Discrimination experiments in cotton-top tamarins (*Saguinus oedipus*) and common marmosets (*Callithrix jacchus*). *Animal Cognition*, 10, 429-437.

Stevens, S.S. (1957). On the psychophysical law. *Psychological Review*, 64 (3), 153 – 181.

Stevens, S.S. (1960). The psychophysics of sensory function. *American Scientist*, 48 (2), 226–253.

Stevens, S.S. (1966). Matching functions between loudness and ten other continua. *Perception & Psychophysics*, *1*, 5-8.

Stoddard, L. T., & McIlvane, W. J. (1989). Generalization after intradimensional discrimination training in 2year-old children. *Journal of Experimental Child Psychology*, 47, 324-334.

Stoddard, L. T., McIlvane, W. J., & de Rose, J. C. (1987). Transferência de controle de estímulo com estudantes deficientes mentais: Modelagem de estímulo, superposição e aprendizagem em uma tentativa. *Psicologia*, 13(3), 13-27.

Sullivan, J. L., Juhasz, B. J., Slattery, T. J., & Barth, H. C. (2011). Adults' number-line estimation strategies: Evidence from eye movements. *Psychonomic Bulletin & Review*, 18, 557–563.

Sutherland, N. S., & Mackintosh, N. J. (1971). *Mechanisms of animal discrimination learning*. New York: Academic Press.

Taatgen, N., & Van Rijn, H. (2010). Nice graphs, good R2, but still a poor fit? How to be more sure your model explains your data. In D. D. Salvucci & G. Gunzelmann (Eds.), Proceedings of the 2010 International Conference on Cognitive Modeling (pp. 247–252). Philadelphia, PA: Drexel.

Tan, L. (2009). The discrimination and representation of relative and absolute number in pigeons and humans. *Unpublished doctoral thesis*, University of Canterbury.

Tan, L., & Grace, R. C. (2012). Human nonverbal discrimination of relative and absolute number. *Learning & Behavior*, 40 (2), 170–179.

Tavolga, W.N. (1969). *Principles of Animal Behavior*. The University of Chicago Press, Chicago, IL, p. 58.

Thomas, R. K., Fowlkes, D., & Vickery, J. D. (1980). Conceptual numerousness judgments by squirrel monkeys. *American Journal of Psychology*, 93, 247-257.

Thompson, C. A., & Opfer, J. E. (2008). Costs and benefits of representational change: Effects of context on age and sex differences in magnitude estimation. *Journal of Experimental Child Psychology*, *101*, 20-51.

Thorndike E.L., & Woodworth, R.S. (1901).The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review*, 8, 247-261. http://psychclassics.yorku.ca/Thorndike/Transfer/transfer1.htm

Tokita, M., & Ishiguchi, A. (2012). Behavioral evidence for format-dependent processes in approximate numerosity representation. *Psychonomic Bulletin & Review*, *19*, 285-293.

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. Psychological Review, 101(1), 80-102.

Van Galen, M.S., & Reistma, P. (2008). Developing access to number magnitude: A study of the SNARC effect in 7- to 9-year-olds. *Journal of Experimental Child Psychology*, 101, 99-113.

VandenBoss, Gary (2006). *APA Dictionary of Psychology*. Washington, DC: American Psychological Association

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

Wearden, J. H. (1992). Temporal generalization in humans. *Journal of Experimental Psychology: Animal Behavior Processes*, 18, 134–144.

Welkowitz, L. A., & Calkins, R. P. (1984). Effects of cognitive and exemplar modeling on field dependence-independence. *Perceptual and Motor Skills*, 58 (2), 439-442.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: the psychophysics of number representation. *Psychological Science*, 10(2), 130-137.

Wildemann, D. G., & Holland, J. G. (1972). Control of a continuous response dimension by a continuous stimulus dimension. *Journal of the Experimental Analysis of Behavior*, 18, 419-434.

Wilson, A., Dehaene, S., Pinel, P., Revkin, S., Cohen, L., & Cohen, D. (2006). Principles underlying the design of ‘‘The Number Race’’, an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions*, 2, 19.

Wilson, A., Revkin, S., Cohen, D., Cohen, L., & Dehaene, S. (2006). An open trial assessment of ‘‘The Number Race’’, an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions*, 2, 20.

Winawer, J., Witthoft, N., Frank, M.C., Wu, L., Wade, A.R., et al. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences, USA*, 104, 7780–7785.

Wood, G., Nuerk, H.C., & Willmes, K. (2006a). Crossed hands and the SNARC effect: A failure to replicate Dehaene, Bossini & Giraux 1993. *Cortex* 42: 1069-1079.

Wood, G., Nuerk, H.C., & Willmes, K. (2006b). Variability of the SNARC effect: Systematic interindividual Differences or just random error? *Cortex, 42:* 1119-1123.

Wynn, K. (1990). Children's understanding of counting. *Cognition*, 36(2) 155-193.

Wynn, K. (1992b). Children's acquisition of number words and the counting system. *Cognitive Psychology*, 24, 220-251.

Zebian, S. (2005). Linkages between Number Concepts, Spatial Thinking, and Directionality of Writing: The SNARC Effect and the REVERSE SNARC Effect in English and Arabic Monoliterates, Biliterates, and Illiterate Arabic Speakers. *Journal of Cognition and Culture*, 5(1), 165 - 190.

Zélanti, P., & Droit-Volet, Sl. (2012). Auditory and visual differences in time perception? An investigation from a developmental perspective with neuropsychological tests. *Journal of Experimental Child Psychology*, 112, 296–311.

Zorzi, M., Priftis, K., & Umiltà, C. (2002). Brain damage: Neglect disrupts the mental number line. *Nature*, *417*, 138–139.

Zorzi, M., Priftis, K., Meneghello, F., Marenzi, R., & Umiltà,C. (2006). The spatial representation of numerical and non-numerical sequences: Evidence from neglect. *Neuropsychologia*, 44, 1061–1067.

APPENDICES

Appendices for Chapter II: Study 1

Preschoolers' categorical vs. adults' continuous mapping patterns in symbolic and nonsymbolic number-to-position tasks

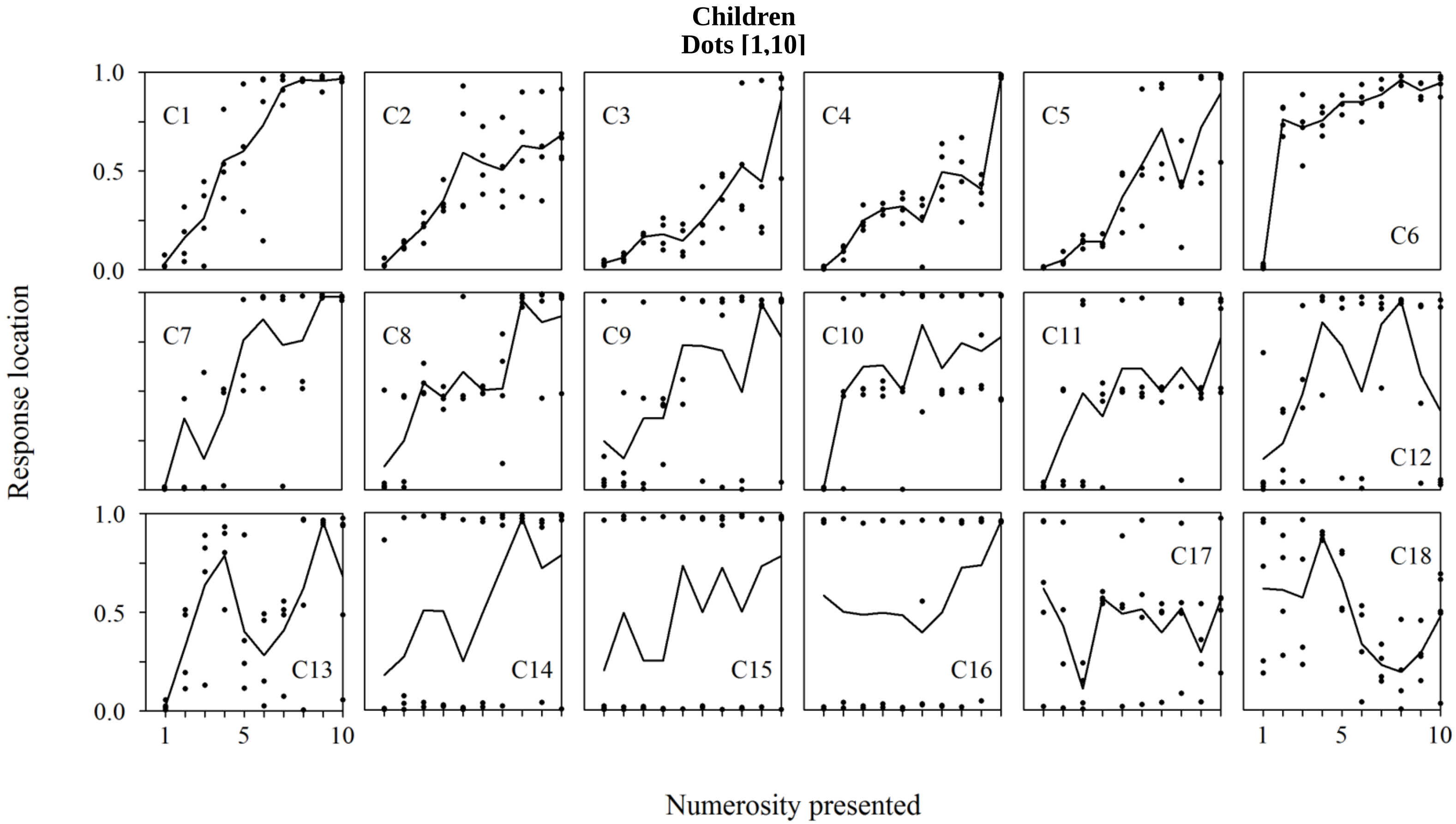
Study 1: Appendix A. Multiple Regression Analyses.

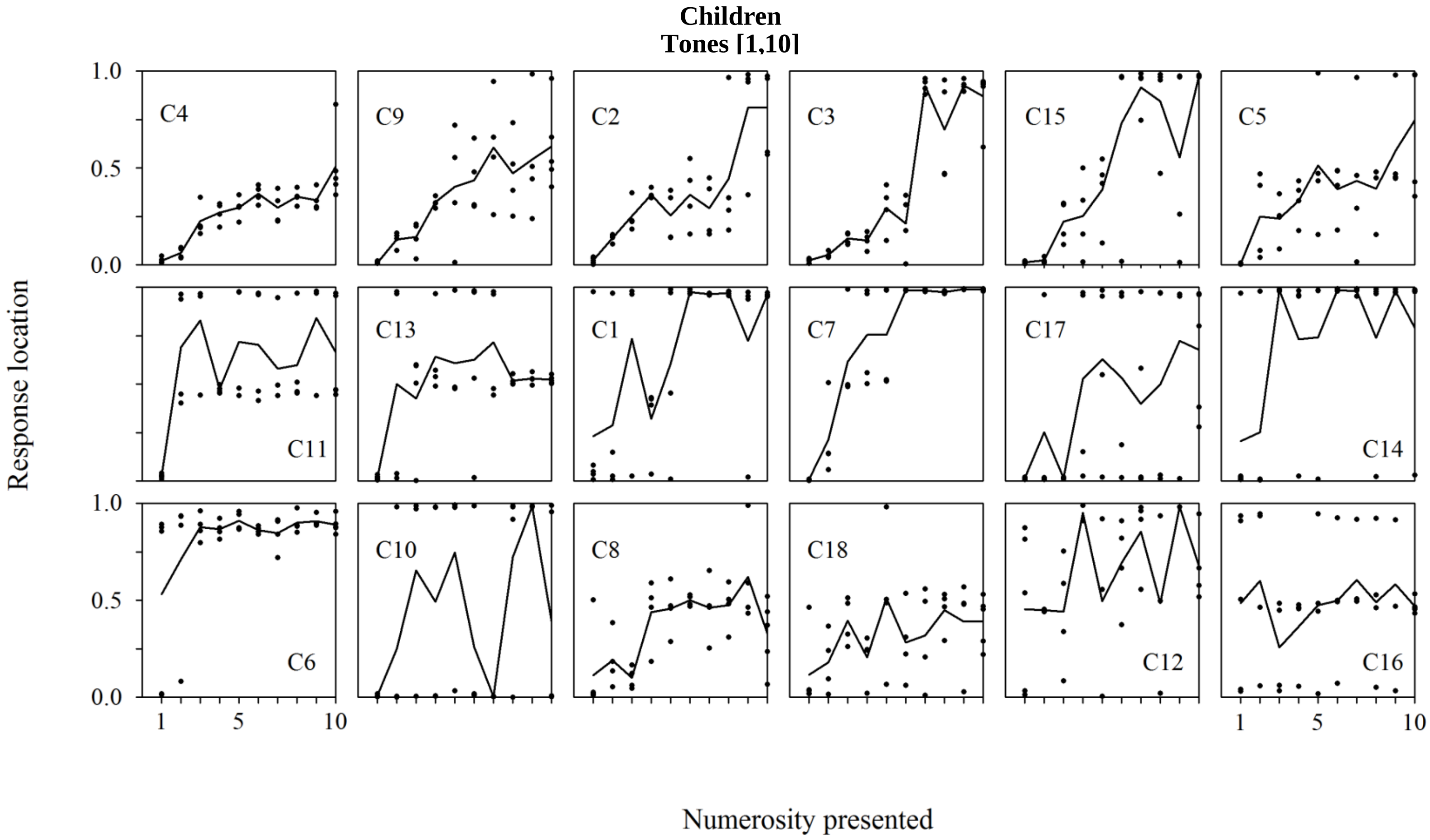
Results of OLS multiple regression analyses, showing the unstandardized weights (plus/minus standard error) of the linear (β_{lin}) and logarithmic (β_{log}) regressors, and its corresponding t ratios and p-values, for each stimulus condition.
In addition to our results (Portuguese adults and preschoolers), we depict the results from similar OLS multiple regression analyses, as reported in Dehaene et al. (2008)'s study with children and adult Munduru and American adults participants, and from the spatial condition in Núñez et al. (2011) study with American adults.

	Stimulus modalities			
	Numerals [1,10]	Tones [1,10]	Dots [1,10]	Dots [10,100]
Pt Adults	$\beta_{lin} = 0.106 \pm 0.004$ t ratio = 28.521 $p < 0.001$	$\beta_{lin} = 0.070 \pm 0.004$ t ratio = 16.725 $p < 0.001$	$\beta_{lin} = 0.104 \pm 0.007$ t ratio = 15.447 $p < 0.001$	$\beta_{lin} = 0.007 \pm 0.001$ t ratio = 6.477 $p < 0.001$
	$\beta_{log} = -0.003 \pm 0.015$ t ratio = -0.216 $p = 0.835$	$\beta_{log} = 0.100 \pm 0.017$ t ratio = 5.782 $p = 0.001$	$\beta_{log} = 0.005 \pm 0.028$ t ratio = 0.167 $p = 0.872$	$\beta_{log} = 0.080 \pm 0.047$ t ratio = 1.712 $p = 0.131$
Pt Children	$\beta_{lin} = 0.034 \pm 0.017$ t ratio = 1.959 $p = 0.091$	$\beta_{lin} = -0.002 \pm 0.008$ t ratio = -0.231 $p = 0.824$	$\beta_{lin} = -0.023 \pm 0.010$ t ratio = 2.352 $p = 0.051$	$\beta_{lin} = 0.034 \pm 0.017$ t ratio = 1.959 $p = 0.091$
	$\beta_{log} = 0.118 \pm 0.072$ t ratio = 1.651 $p = 0.143$	$\beta_{log} = 0.256 \pm 0.033$ t ratio = 7.704 $p < 0.001$	$\beta_{log} = 0.162 \pm 0.040$ t ratio = 4.009 $p = 0.005$	$\beta_{log} = 0.118 \pm 0.072$ t ratio = 1.651 $p = 0.143$
Núñez et al. (2011)				
USA adults	$\beta_{lin} = 0.093 \pm 0.006$ t ratio = 16.340 $p < 0.001$	$\beta_{lin} = 0.071 \pm 0.007$ t ratio = 11.400 $p < 0.001$	$\beta_{lin} = 0.084 \pm 0.010$ t ratio = 8.890 $p < 0.001$	$\beta_{lin} = 0.002 \pm 0.001$ t ratio = 1.580 $p = 0.158$
	$\beta_{log} = 0.106 \pm 0.054$ t ratio = 1.950 $p = 0.092$	$\beta_{log} = 0.238 \pm 0.060$ t ratio = 3.980 $p = 0.005$	$\beta_{log} = 0.170 \pm 0.090$ t ratio = 1.880 $p = 0.102$	$\beta_{log} = 0.742 \pm 0.096$ t ratio = 7.810 $p < 0.001$
Dehaene et al. (2008)				
USA adults	$\beta_{lin} = 0.107 \pm 0.005$ t ratio = 19.844 $p < 0.001$	$\beta_{lin} = 0.023 \pm 0.015$ t ratio = 1.514 $p = 0.174$	$\beta_{lin} = 0.103 \pm 0.007$ t ratio = 14.342 $p < 0.001$	$\beta_{lin} = 0.030 \pm 0.015$ t ratio = 1.982 $p = 0.088$
	$\beta_{log} = 0.009 \pm 0.022$ t ratio = 0.406 $p = 0.697$	$\beta_{log} = 0.244 \pm 0.063$ t ratio = 3.857 $p = 0.006$	$\beta_{log} = 0.032 \pm 0.029$ t ratio = 1.146 $p = 0.290$	$\beta_{log} = 0.343 \pm 0.066$ t ratio = 5.144 $p = 0.001$
Mundurucu	$\beta_{lin} = 0.018 \pm 0.015$ t ratio = 1.184 $p = 0.290$	$\beta_{lin} = 0.014 \pm 0.009$ t ratio = 1.532 $p = 0.169$	$\beta_{lin} = 0.023 \pm 0.015$ t ratio = 1.514 $p = 0.174$	$\beta_{lin} = -0.027 \pm 0.031$ t ratio = 0.887 $p = 0.404$
	$\beta_{log} = 0.207 \pm 0.064$ t ratio = 3.145 $p = 0.024$	$\beta_{log} = 0.122 \pm 0.038$ t ratio = 3.240 $p = 0.015$	$\beta_{log} = 0.248 \pm 0.064$ t ratio = 3.857 $p = 0.006$	$\beta_{log} = 0.507 \pm 0.127$ t ratio = 3.931 $p = 0.005$

Study 1: Appendix B. Individual scatterplots of the Number-to-Position tasks.

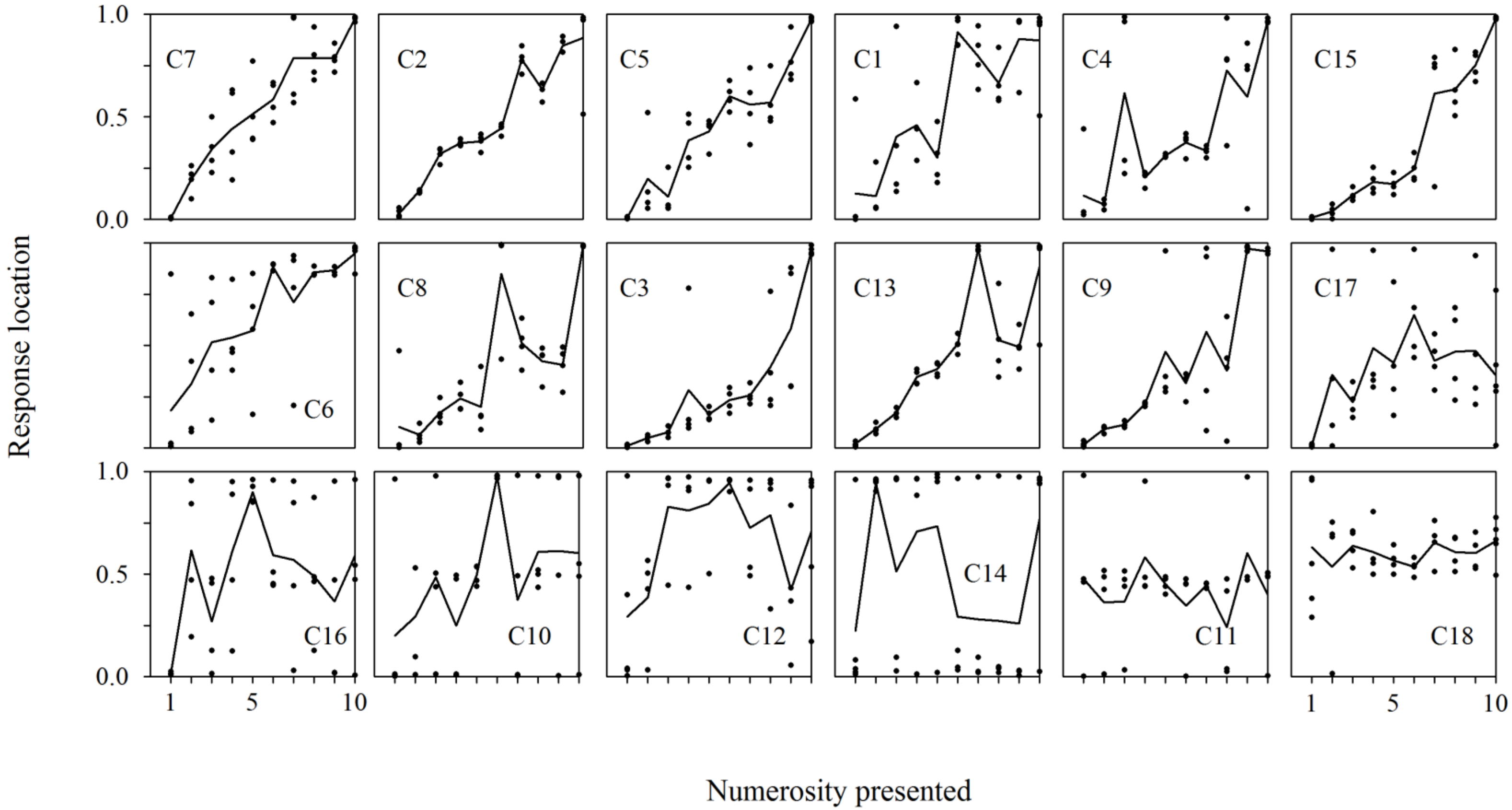
Each dot is one trial's response location and the line depicts mean location.
Graphs were arranged according to the increasing use of the extent of the response bar, namely, from a "continuous", to a "tri-categorical" and "bi-categorical" performance. Whenever classification proved doubtful, we carried a k-means cluster analysis to determine the suitable cut-off (number of clusters). The last graph cases on each condition are the participants whose mean location did not increase as a function of numerosity.


Children
Dots [1,10]
Response location
Numerosity presented
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12
C13
C14
C15
C16
C17
C18
1.0
0.5
0.0
1
5
10

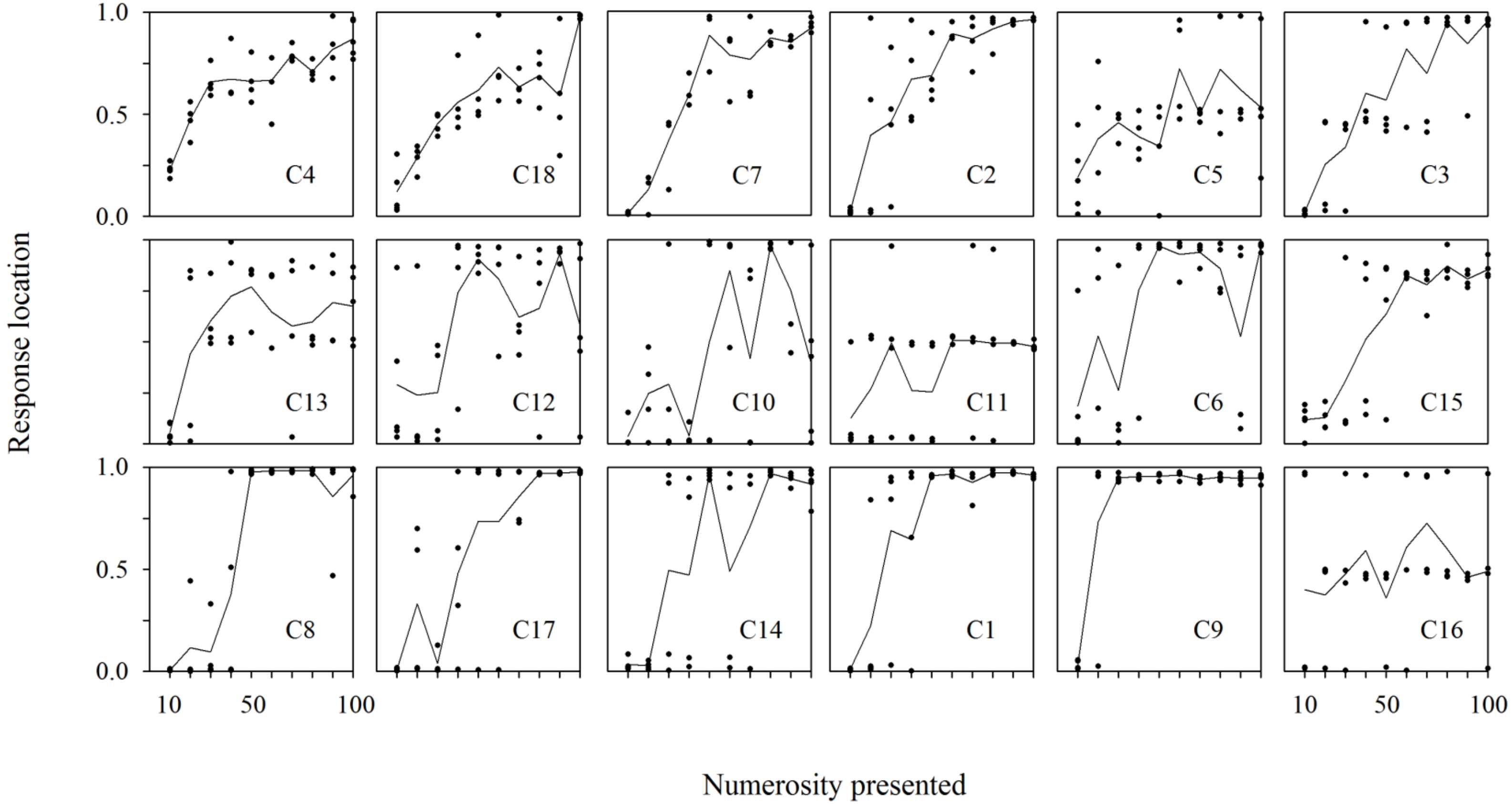


Children
Tones [1,10]
Response location
1.0
0.5
0.0
C4
C9
C2
C3
C15
C5
C11
C13
C1
C7
C17
C14
C6
C10
C8
C18
C12
C16
1
5
10
Numerosity presented

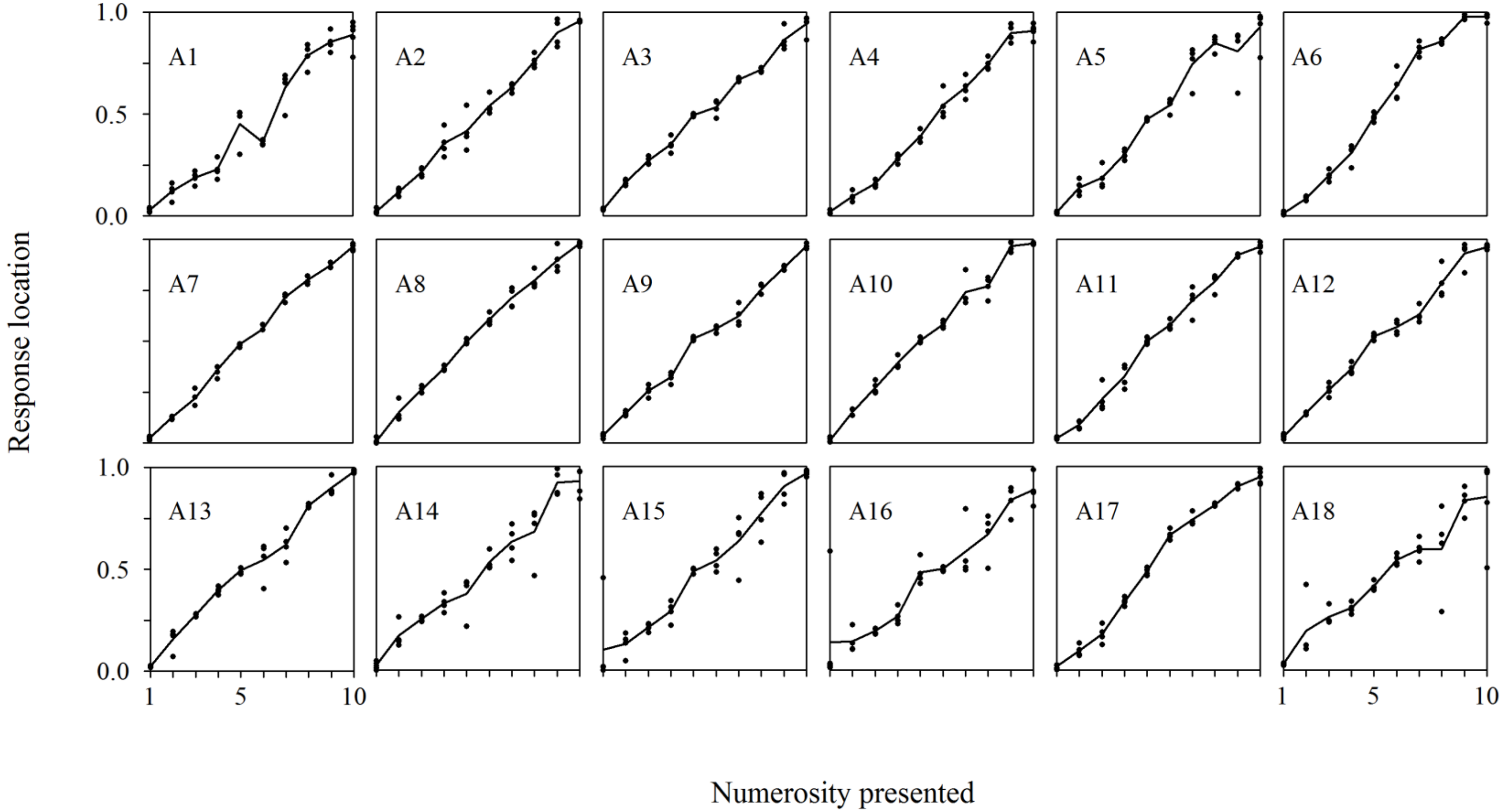
Children
Verbal [1,10]

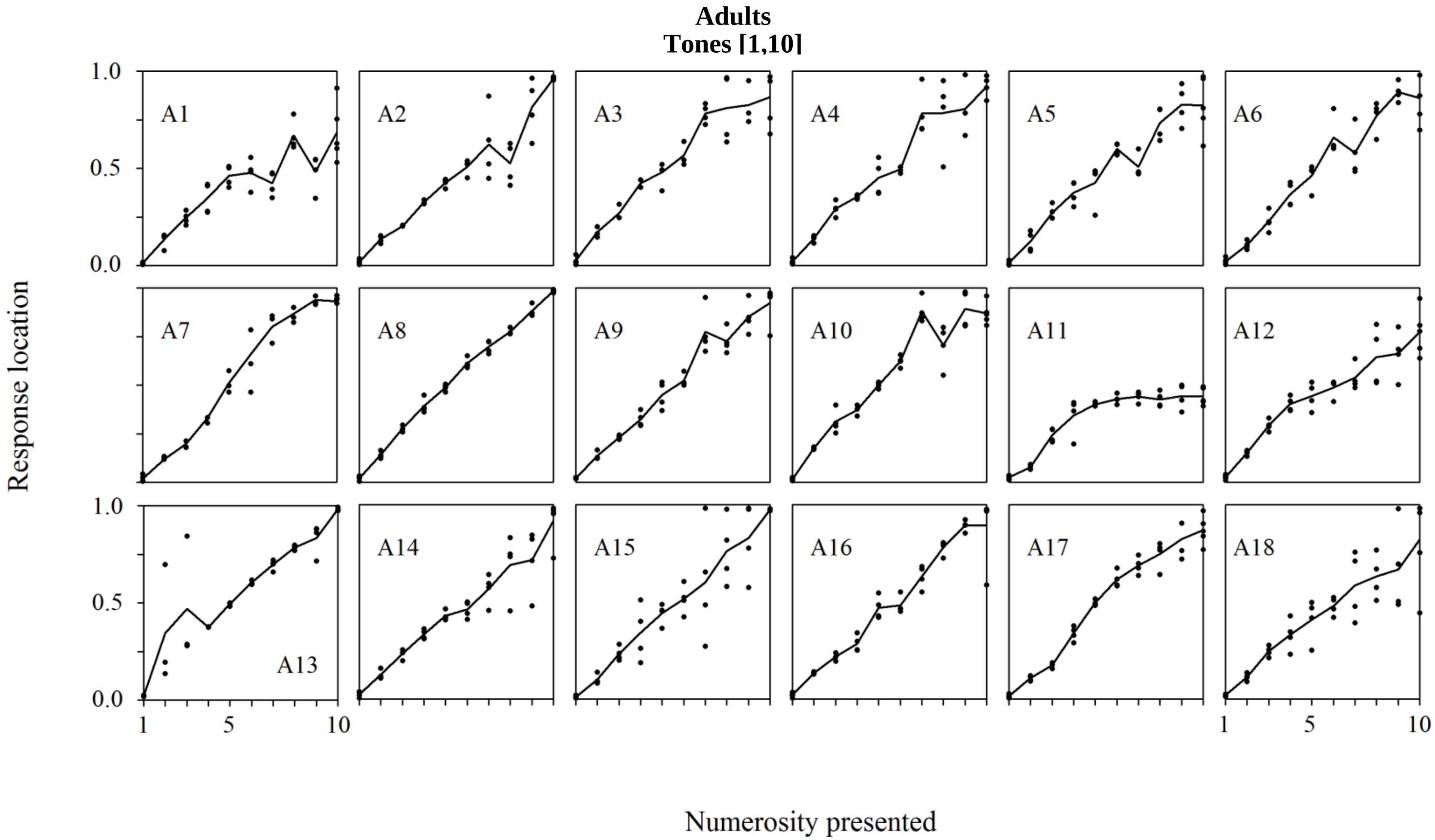


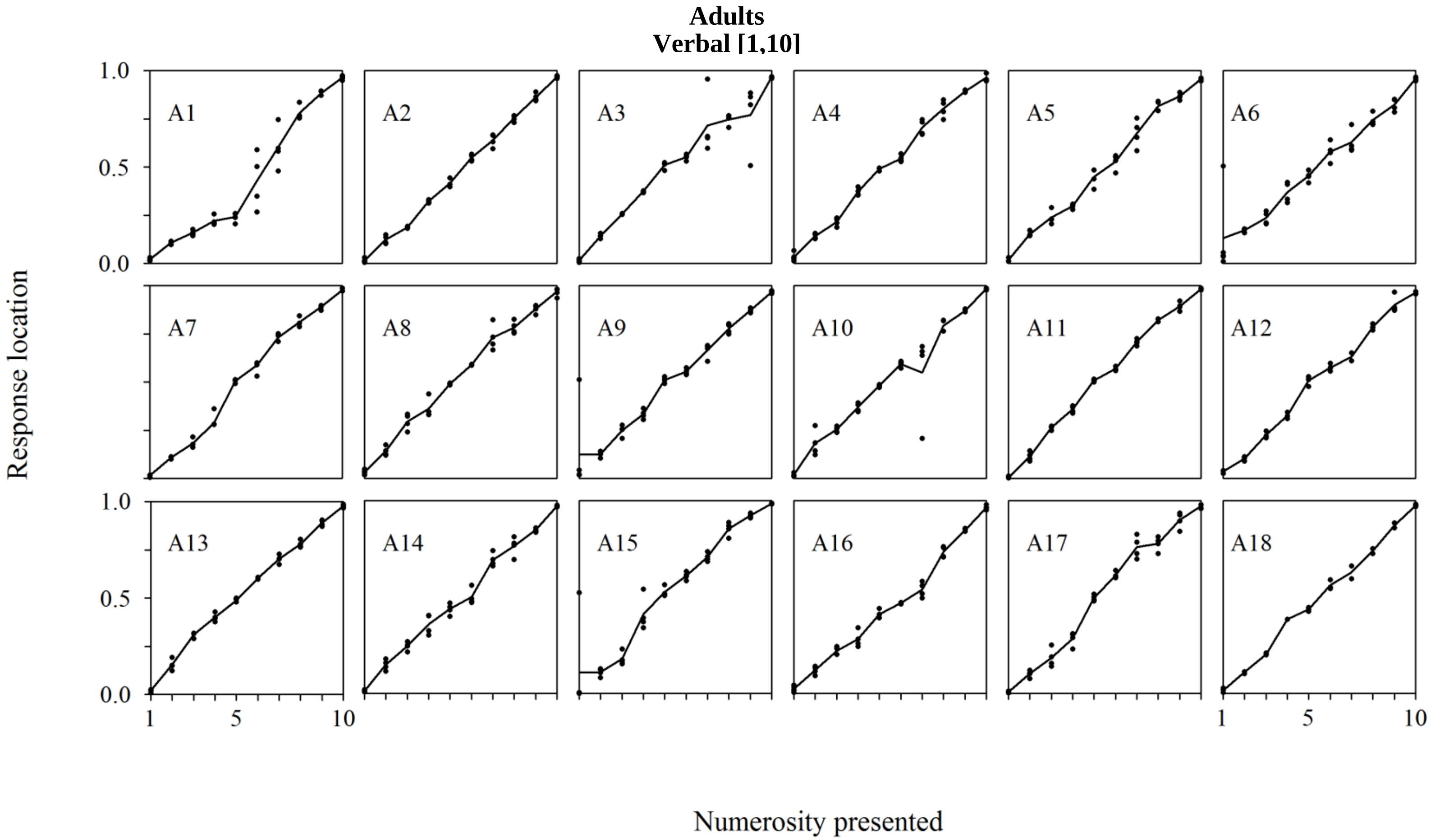
Children
Dots [10,100]

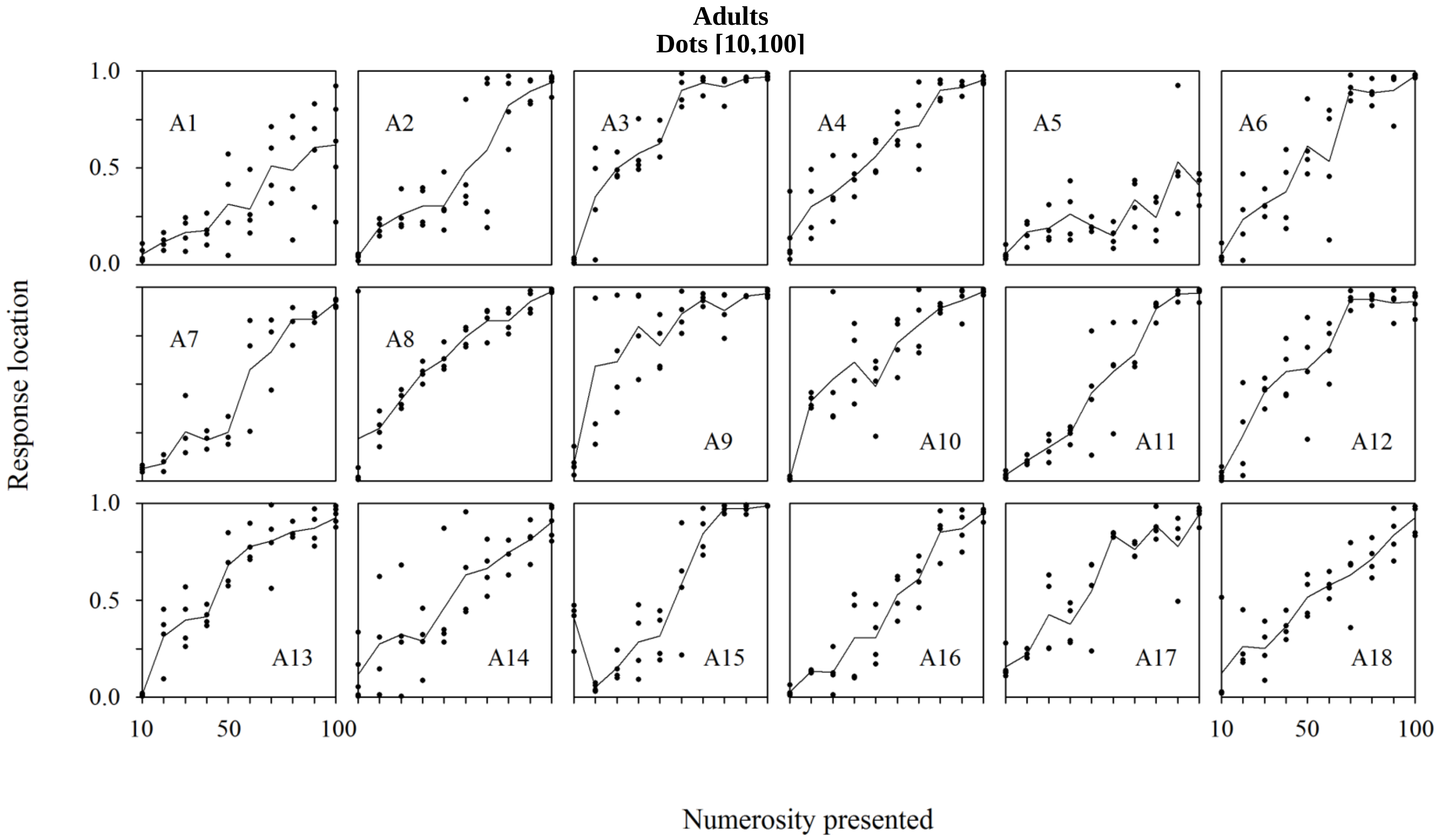


Adults
Dots [1,10]




Adults
Tones [1,10]
Response location
Numerosity presented
A1
A2
A3
A4
A5
A6
A7
A8
A9
A10
A11
A12
A13
A14
A15
A16
A17
A18
1.0
0.5
0.0
1
5
10


Adults
Verbal [1,10]
Response location
1.0
0.5
0.0
A1
A2
A3
A4
A5
A6
A7
A8
A9
A10
A11
A12
A13
A14
A15
A16
A17
A18
1
5
10
Numerosity presented



Adults
Dots [10,100]
Response location
1.0
0.5
0.0
A1
A2
A3
A4
A5
A6
A7
A8
A9
A10
A11
A12
A13
A14
A15
A16
A17
A18
10
50
100
Numerosity presented

Study 1: Appendix C. Individual normalized entropy scores.

Entropy was estimated based on the frequency distribution of the response locations, according to the formula $-\frac{\sum_{i=1}^{10}(p_i \times log_2 p_i)}{log_2 10}$.

Children					**Adults**				
	Dots [1,10]	Tones [1,10]	Numerals [1,10]	Dots [10,100]		Dots [1,10]	Tones [1,10]	Numerals [1,10]	Dots [10,100]
C1	.75	.43	.92	.39	**A1**	.97	.90	.89	.91
C2	.96	.77	.94	.73	**A2**	.98	.96	.94	.86
C3	.78	.77	.76	.49	**A3**	1.0	.97	.95	.76
C4	.83	.68	.79	.83	**A4**	.97	.96	.98	.92
C5	.77	.70	.89	.79	**A5**	.95	.95	.97	.70
C6	.66	.43	.81	.56	**A6**	.91	.96	.97	.91
C7	.51	.50	.95	.80	**A7**	.96	.94	.92	.91
C8	.69	.81	.81	.48	**A8**	.98	.99	.98	.94
C9	.57	.92	.73	.18	**A9**	.92	.97	.96	.76
C10	.66	.30	.59	.68	**A10**	.96	.96	.97	.86
C11	.60	.46	.43	.54	**A11**	.95	.58	.91	.84
C12	.59	.81	.64	.72	**A12**	.92	.94	.95	.84
C13	.85	.61	.83	.71	**A13**	.96	.95	.96	.89
C14	.34	.25	.38	.40	**A14**	.96	.94	.95	.94
C15	.30	.71	.85	.66	**A15**	.96	.90	.89	.86
C16	.33	.57	.73	.50	**A16**	.92	.96	.93	.93
C17	.81	.56	.91	.55	**A17**	.94	.95	.95	.83
C18	.97	.71	.75	.95	**A18**	.97	.92	.96	.98
avg	.66	.61	.76	.61	avg	.95	.93	.95	.87
SD	.19	.18	.16	.18	SD	.02	.09	.03	.07

Appendices for Chapter III: Study 2

Control of a response continuum by the numerical stimulus continuum: the effects of pre-training on a non-numerical continuum

Study 2: Appendix A. Verbal instructions that preceded the Number-to-Position computerized sessions

Table. Script of the instructions given to the participant (in Portuguese, and their translation into English. The instructions were given by the experimenter (Ɛ) when: (2.0) the response bar was first presented to the participant; (2.1) before the participant started the training with the two anchor numerosities; and, finally, (2.2) before the participant was tested with both the anchors and novel, intermediate numerosities.

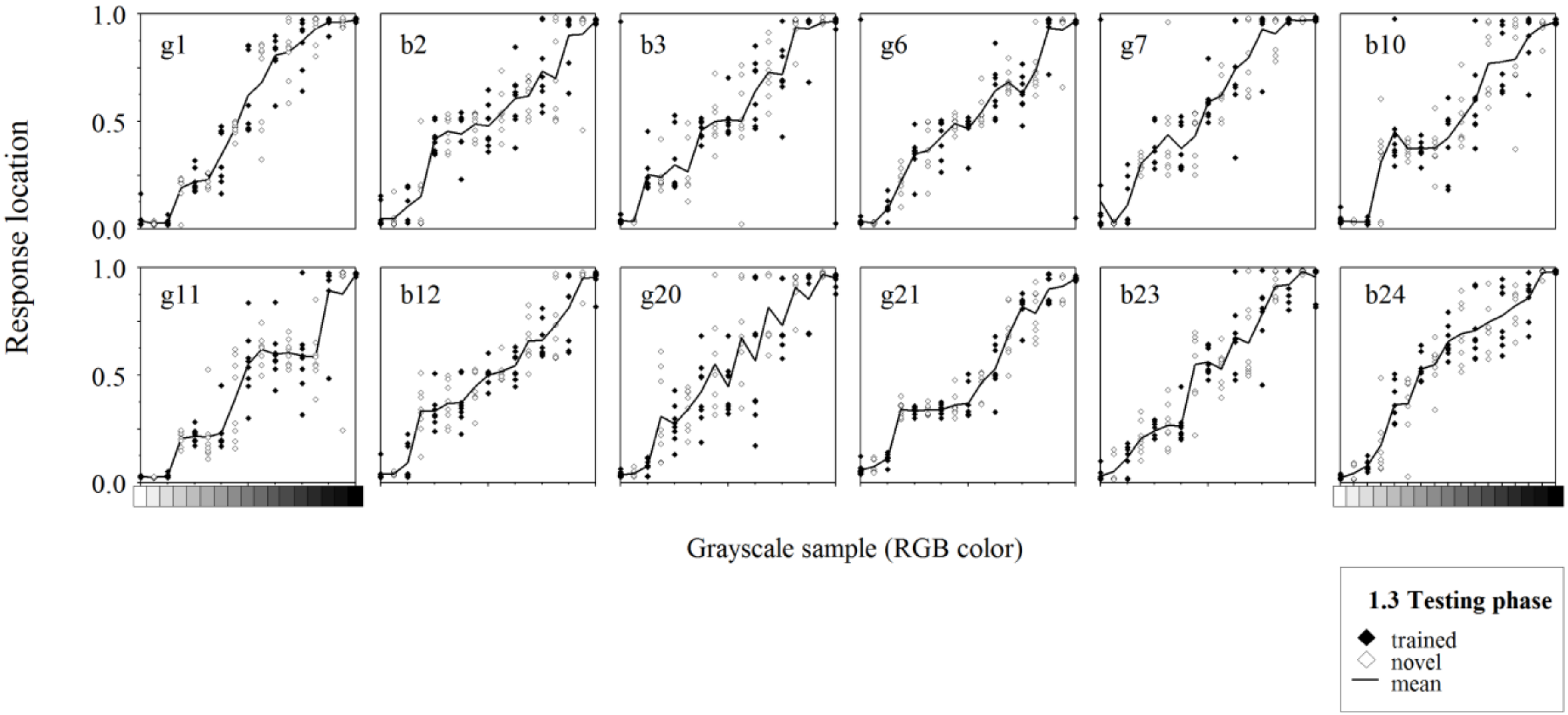
Experimental Phase	Instructions in Portuguese	Translation to English
2.0 **Response Bar**	(Experimentadora (Ɛ) aponta para barra amarela) "*Estás a ver esta barra? De que cor é?*" (criança response; Ɛ corrige-a). "*A barra amarela é um caminho. Este caminho começa aqui*" (Ɛ aponta para posição mais à esquerda). "*Queres mostrar-me tu onde começa o caminho?*" (criança aponta; Ɛ corrige-a). "*Vamos andar mais no caminho?*" (Ɛ segura na mão da criança e movimenta-a de modo a que o dedo indicador toque na barra). "*Vamos andando, e andando mais…*" (o dedo toca em mais quatro posições, da esquerda para a direita, até chegar ao extremo direito da barra). "*Chegámos ao fim do caminho.*" "*Agora, mostra-me tu onde começa o caminho e depois anda no caminho até chegares ao fim.*" (A criança deve por tocar no extremo esquerdo e ir tocando a barra em sucessivas posições à direita até parar no extremo direito; Caso contrário, Ɛ corrige-a).	(Experimenter (Ɛ) points to the yellow bar) "*Do you see this bar? What color is it?*" (child answers; Ɛ corrects his/her). "*The yellow bar is a path. This path starts here*" (Ɛ points to the leftmost position). "*Do you want to show me where does the path start?*" (child points; Ɛ corrects). "*Shall we walk further into the path?*" (Ɛ holds the child's hand and moves it so that the index finger touches the bar). "*We are walking, and walking more…*" (the finger touches four more positions, in the left-to-right orientation, until it reaches the bars' rightmost position). "*We have arrived at the path's end*." "*Now, you show me where does the path start and then walk the path until you reach its end.*" (The child should touch the leftmost position then touch the bar progressively to the right until it stops at the rightmost position; Otherwise, Ɛ corrects).
2.1 **Training** **2 anchor numerosities**	"*Este caminho é onde ficam os números. Os números moram no caminho*." "*O número mais pequenino, que são as mais pouquinhas bolas fica/mora aqui*," (Ɛ aponta para extremo esquerdo) "*e o número maior fica ali*." (Ɛ aponta para extremo direito) "*Quanto maior o número, quando mais bolinhas, mais temos de andar no caminho*." (Ɛ move dedo ao longo da barra, da esquerda para a direita, tocando em 5 posições equidistantes). "*Mas por agora vamos ver só o número mais pequeno possível e o número maior possível de bolinhas, e vamos ensiná-los onde ficam no caminho*".	"*This path is where the numbers stay. The numbers reside in the path*." "*The smallest number, which is the fewest dots, stays/belongs here*," (Ɛ points to leftmost position) "*and the largest number stays there*." (Ɛ points to rightmost position). "*The larger the number, the more dots there are, the more we have to walk along the path*." (Ɛ moves her finger along the bar, from left to right, touching at 5 equidistant positions). "*But for now we will only see the fewest and the largest possible number of dots, and teach them where they are to stay along the path*.

	● Instruções durante Feedback e Procedimento de Correcção: Âncora com menor numerosidade: "*Uma vez que este é o menor número possível de bolinhas, ele fica no início do caminho, não precisa de andar nada*." (Ɛ aponta e toca no extremo esquerdo) Âncora com maior numerosidade: "*Uma vez que este é o maior número possível de bolinhas, ele tem de andar tudo isto até ao final do caminho*." (Ɛ desloca o dedo ao longo da barra até parar e tocá-la no extremo direito)	● Instructions during Feedback and Correction Procedure: Smallest anchor numerosity: "*Because this is the fewest possible number of dots, it stays at the beginning of the path, does not need to walk at all*." (Ɛ points and touches at the bars' leftmost position) Largest anchor numerosity: "*Because this is the largest possible number of dots, it has to walk all this way up until the end of the path*." (Ɛ moves finger along the bar until it stops and touches it at its rightmost position)
2.2 **Testing** **2 anchors plus 7 intermediate numerosities**	"*A partir de agora, já não vamos ver só o menor e o maior número possível de bolinhas. Vamos ver também outros números de bolinhas.* *Quando vires outro número de bolinhas vais ter de decidir até onde é que elas devem ficar ao longo do caminho. Quando vires um número novo deves lembrar-te que quanto mais bolinhas, mais elas andam ao longo do caminho. Quanto mais o número aumenta, mais ele tem de andar no caminho*." (enquanto diz isto, Ɛ move o seu dedo ao longo da barra, da esquerda para a direita, tocando a barra em cinco posições).	"*From now on, we will not just see only the fewest possible and largest possible number of dots. We will also see other numbers of dots.* *Whenever you see other number of dots you will have to decide how further they must stay along the path. When you see a novel number you must remember that the more the dots, the further they must walk along the path. The bigger the number, the more it has to walk along the path*." (while saying this, Ɛ moved her finger along the path, from left-to-right, at five locations along the bar).

Study 2: Appendix B. Individual scatterplots of Brightness-to-position tasks: *1.3 Testing of trained and novel intermediate values***.**

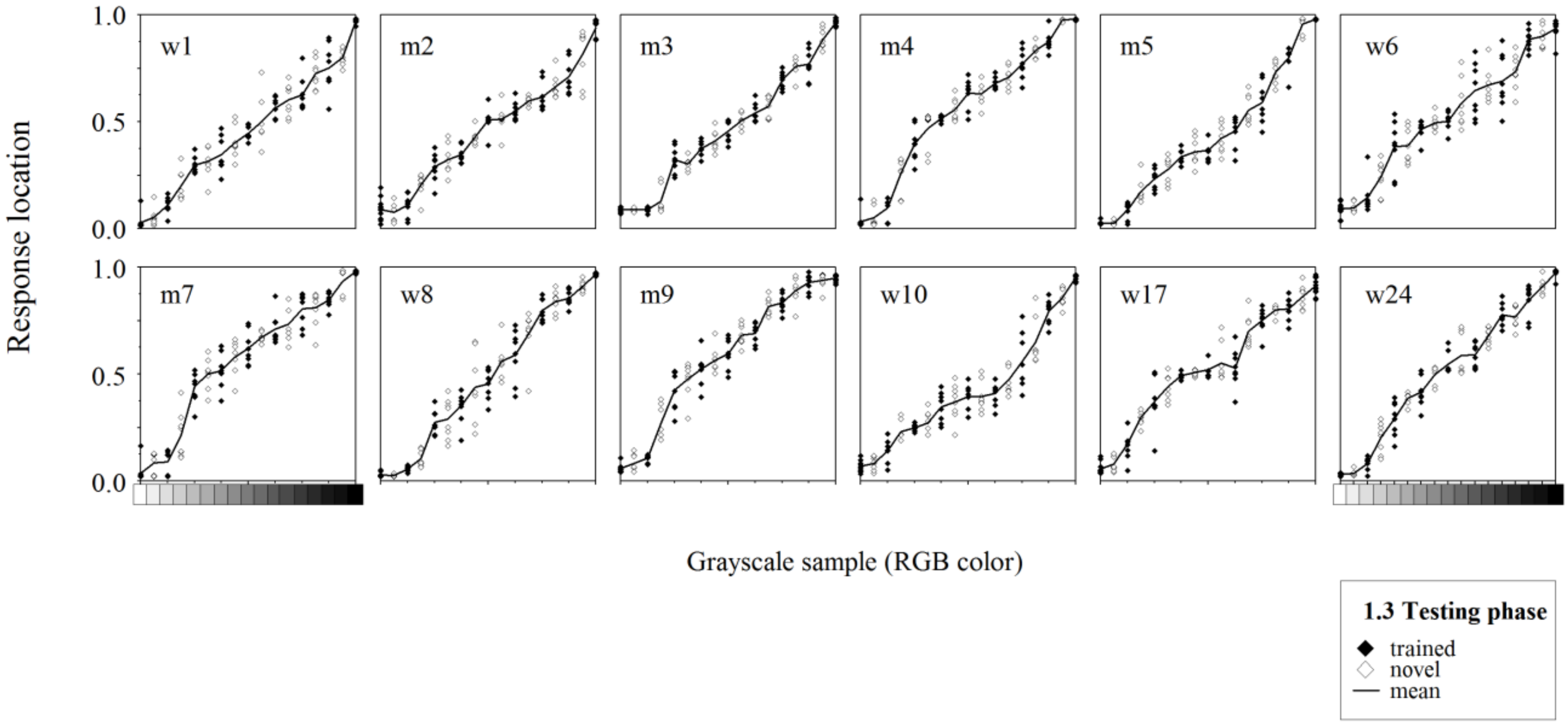
Each data point is one trial's response location. The line depicts mean location. Filled diamonds represent the response location for previously trained sample hues, the white diamonds the responses for the novel, interpolated hues.

Children



Response location
1.0
0.5
0.0
g1
b2
b3
g6
g7
b10
g11
b12
g20
g21
b23
b24
Grayscale sample (RGB color)
1.3 Testing phase
trained
novel
mean

Adults


1.0
0.5
0.0
Response location
w1
m2
m3
m4
m5
w6
m7
w8
m9
w10
w17
w24
Grayscale sample (RGB color)
1.3 Testing phase
trained
novel
mean

Study 2: Appendix C. Multiple Regression Analyses (group data)**.**

Table 1. Results of OLS multiple regression analyses, showing the unstandardized weights (plus/minus standard error) of the linear (β_{lin}) and logarithmic (β_{log}) regressors, and its corresponding t ratios and p-values (with df = 7), for each stimulus condition.

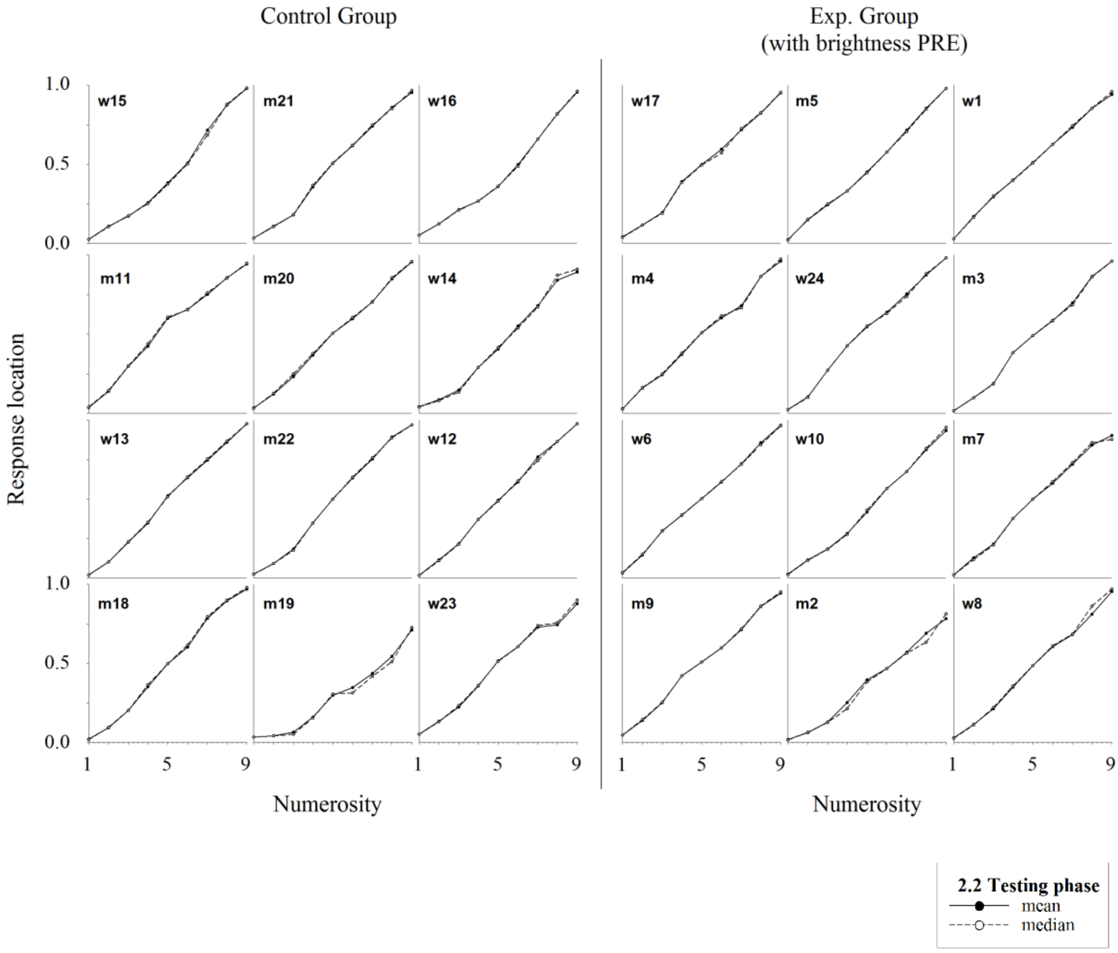
	Dots [1,9]		Dots [10,90]	
	Mean	Median	Mean	Median
Adults				
Ctrl.	β_{lin} = 0.136 ± 0.006 t ratio = 22.692 $p < 0.001$	β_{lin} = 0.136 ± 0.008 t ratio = 16.326 $p < 0.001$	β_{lin} = 0.009 ± 0.001 t ratio = 8.827 $p < 0.001$	β_{lin} = 0.010 ± 0.001 t ratio = 9.159 $p < 0.001$
	β_{log} = -0.173 ± 0.052 t ratio = -3.299 $p = 0.016$	β_{log} = -0.134 ± 0.073 t ratio = -1.826 $p = 0.118$	β_{log} = 0.217 ± 0.085 t ratio = 2.553 $p = 0.043$	β_{log} = 0.111 ± 0.100 t ratio = 1.113 $p = 0.308$
Exp.	β_{lin} = 0.121 ± 0.004 t ratio = 27.350 $p < 0.001$	β_{lin} = 0.124 ± 0.006 t ratio = 20.963 $p < 0.001$	β_{lin} = 0.006 ± 0.001 t ratio = 12.418 $p < 0.001$	β_{lin} = 0.008 ± 0.001 t ratio = 6.551 $p = 0.001$
	β_{log} =- 0.052 ± 0.039 t ratio = -1.332 $p = 0.231$	β_{log} =- 0.053 ± 0.052 t ratio = -1.027 $p = 0.344$	β_{log} = 0.402 ± 0.046 t ratio = 8.829 $p < 0.001$	β_{log} = 0.323 ± 0.111 t ratio = 2.916 $p = 0.027$
Children				
Ctrl.	β_{lin} = 0.104 ± 0.011 t ratio = 9.596 $p < 0.001$	β_{lin} = 0.164 ± 0.019 t ratio = 8.638 $p < 0.001$	β_{lin} = 0.004 ± 0.003 t ratio = 1.227 $p = 0.266$	β_{lin} = 0.004 ± 0.008 t ratio = 0.511 $p = 0.628$
	β_{log} = -0.117 ± 0.095 t ratio = -1.238 $p = 0.262$	β_{log} = -0.559 ± 0.166 t ratio = -3.365 $p = 0.015$	β_{log} = 0.675 ± 0.269 t ratio = 2.506 $p = 0.046$	β_{log} = 0.905 ± 0.679 t ratio = 1.332 $p = 0.231$
Exp.	β_{lin} = 0.118 ± 0.006 t ratio = 20.882 $p < 0.001$	β_{lin} = 0.148 ± 0.008 t ratio = 18.216 $p < 0.001$	β_{lin} = 0.003 ± 0.001 t ratio = 2.832 $p = 0.030$	β_{lin} = 0.005 ± 0.003 t ratio = 1.467 $p = 0.193$
	β_{log} =- 0.093 ± 0.050 t ratio = -1.878 $p = 0.109$	β_{log} = -0.234 ± 0.071 t ratio = -3.294 $p = 0.017$	β_{log} = 0.718 ± 0.096 t ratio = 7.502 $p < 0.001$	β_{log} = 0.748 ± 0.271 t ratio = 2.762 $p = 0.033$

Study 2: Appendix D. Mean and Median scatterplots of Number-to-position tasks (2.2 Testing of 2 anchors and 7 intermediate numerosities).

Graphs are separated according to the experimental group and ordered by participants' age. Black circles connected by the solid line represent the mean, and white circles connected by the dashed line the median response location.

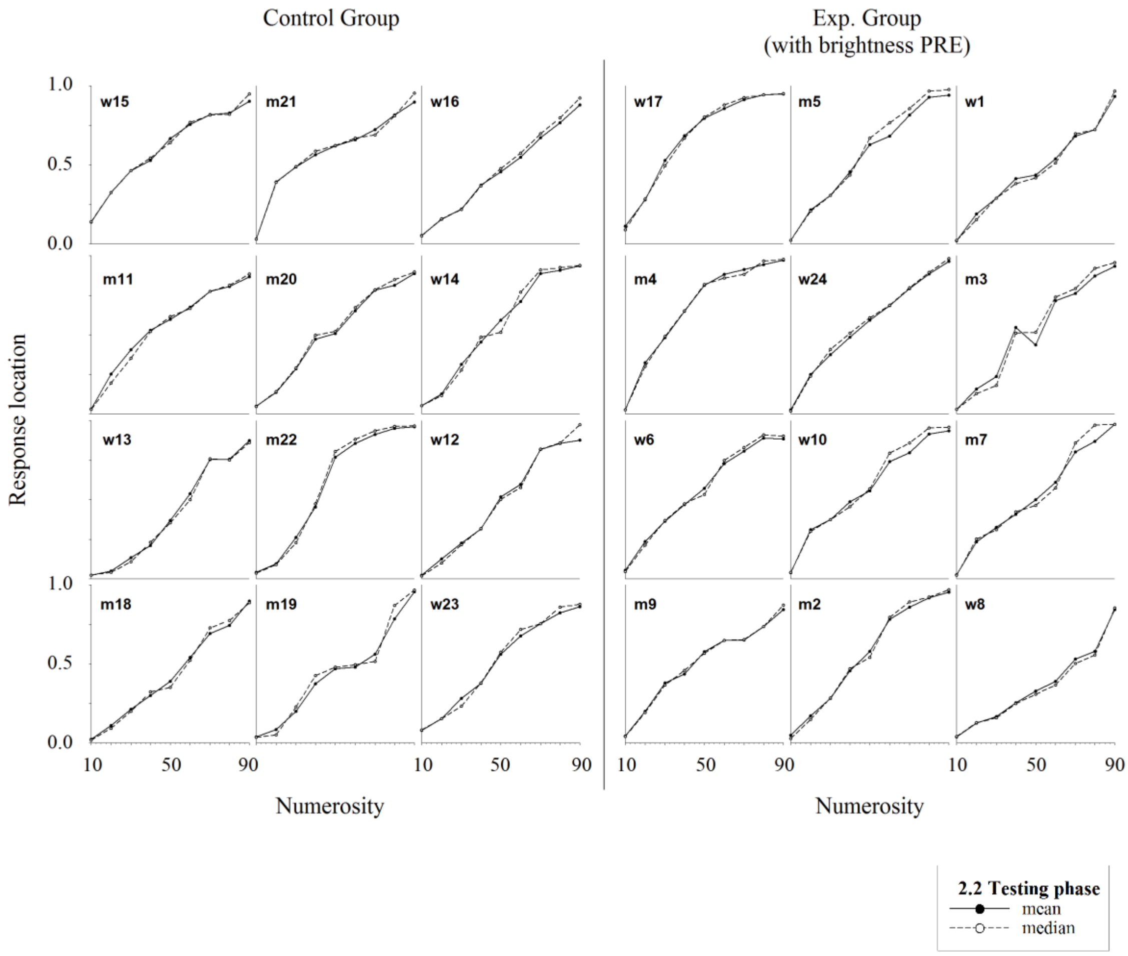
Adults

Dots [1,9]



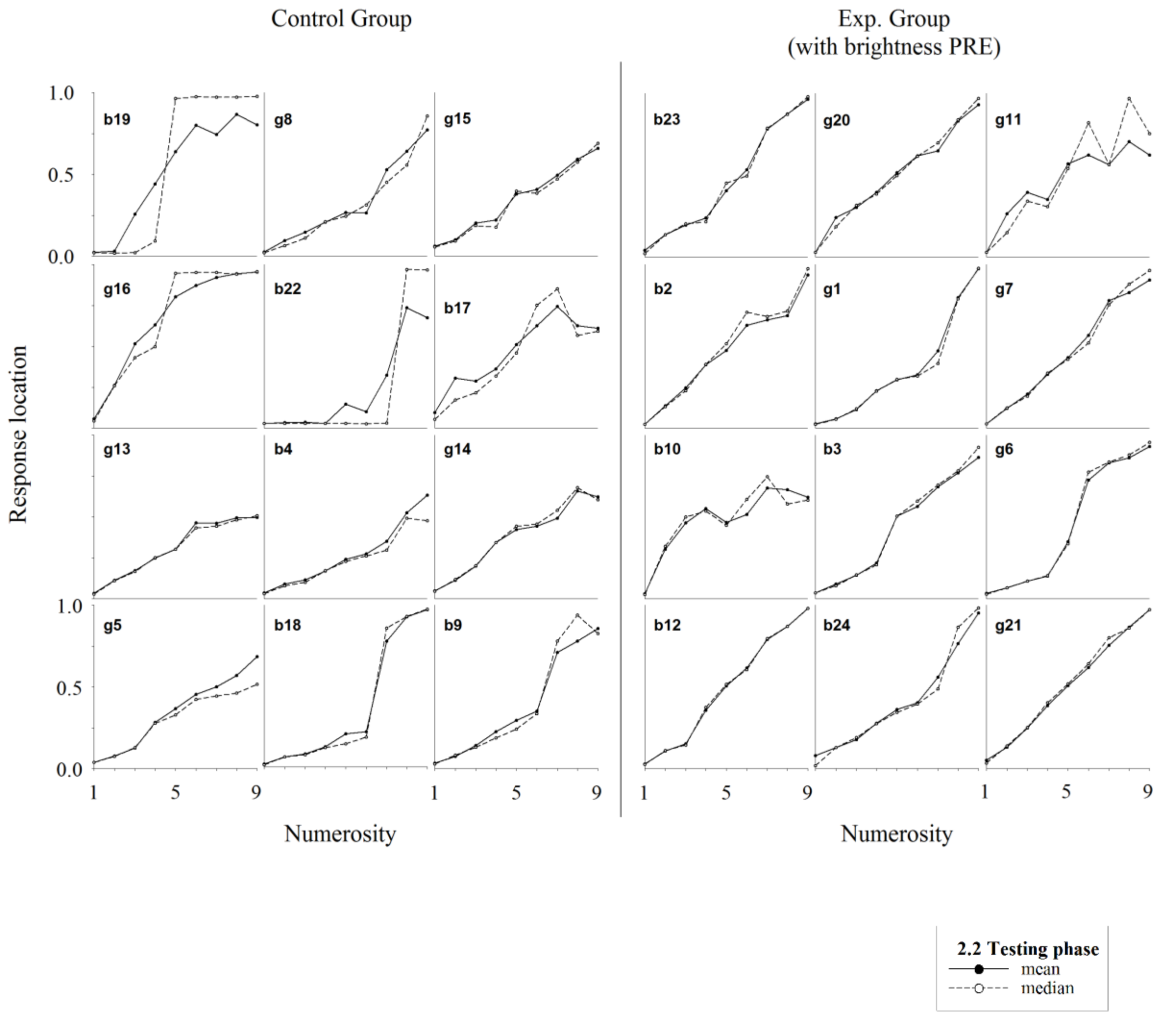
Adults

Dots [10,90]



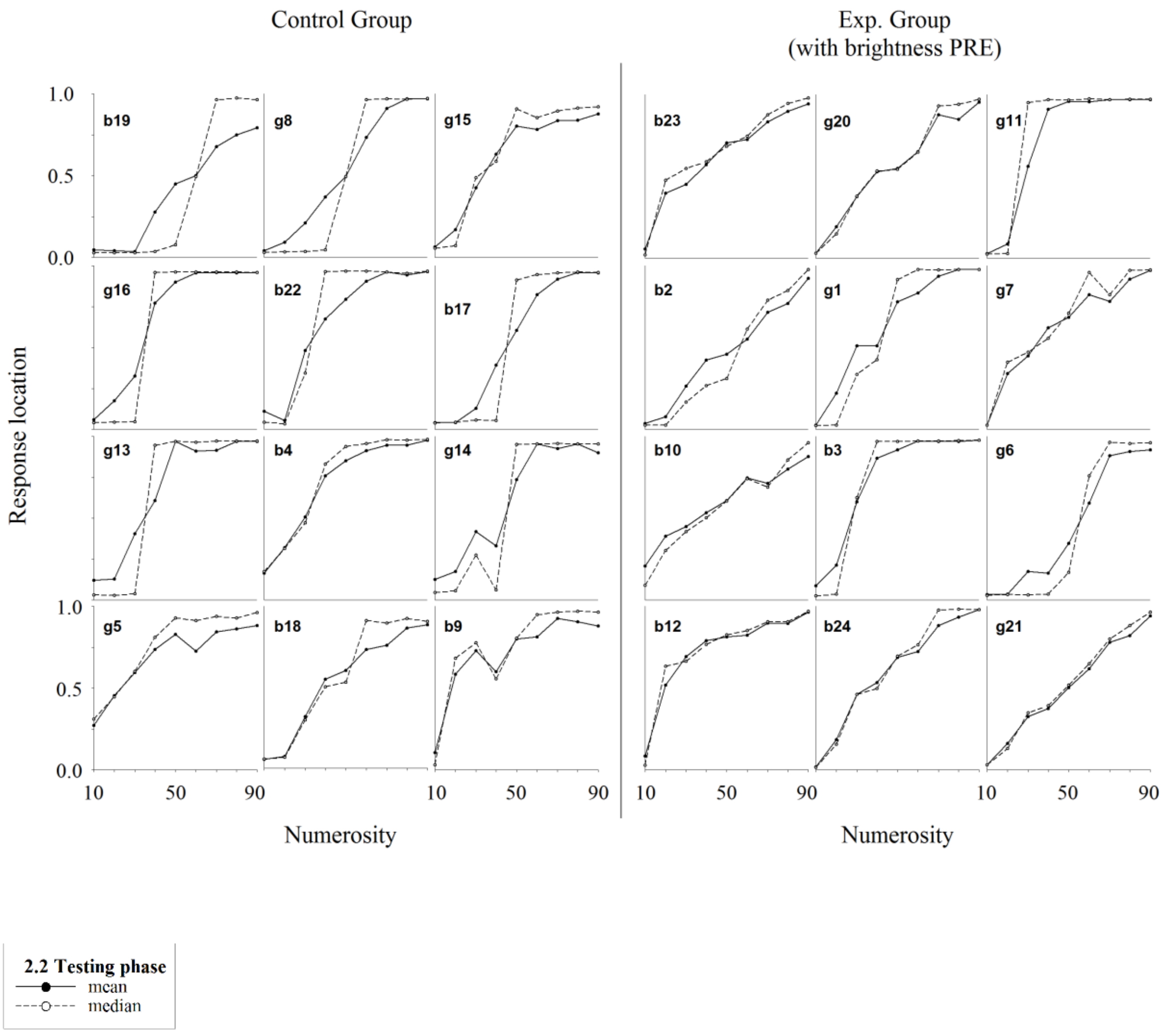
Children

Dots [1,9]



Children

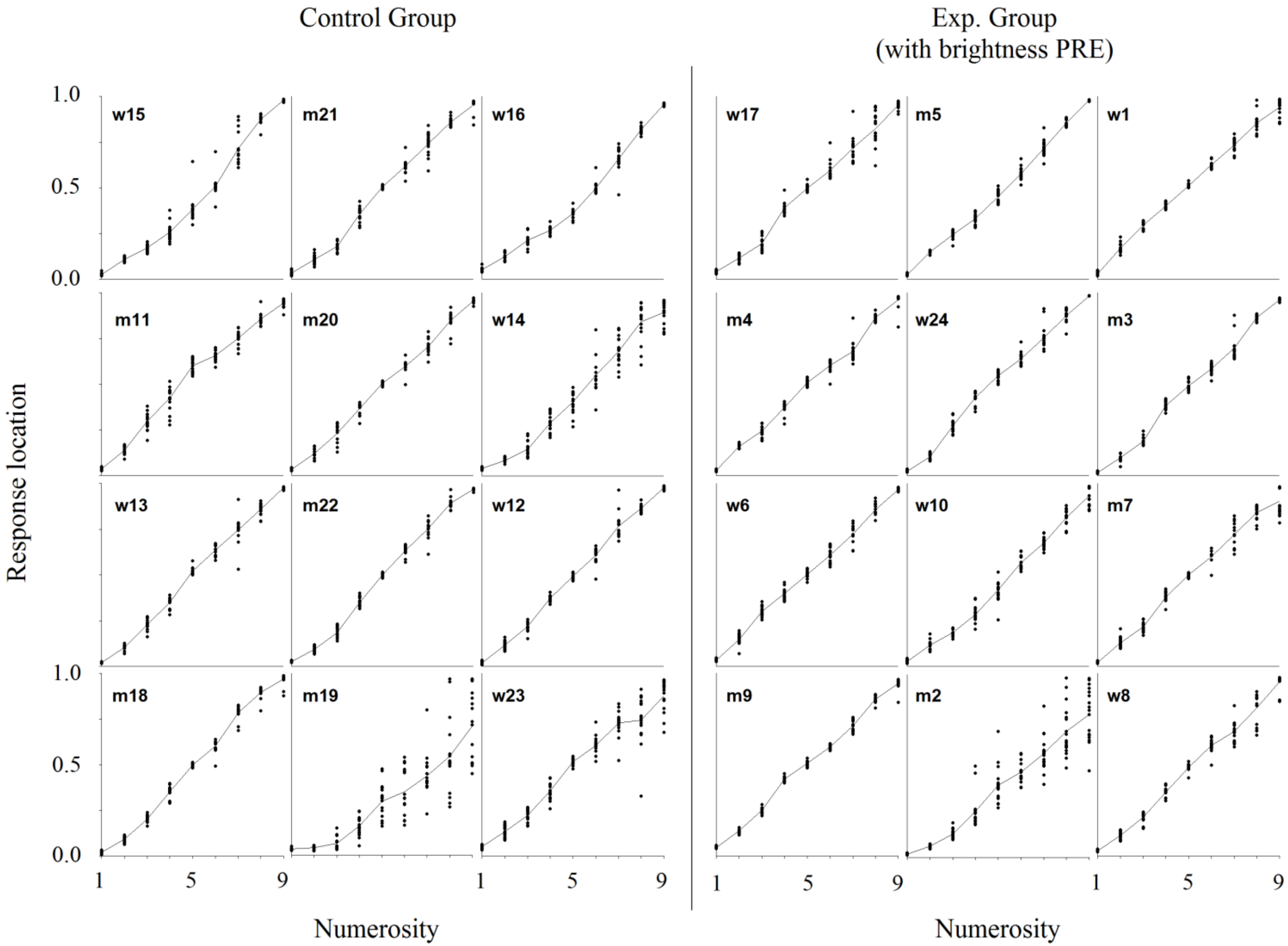
Dots [10,90]



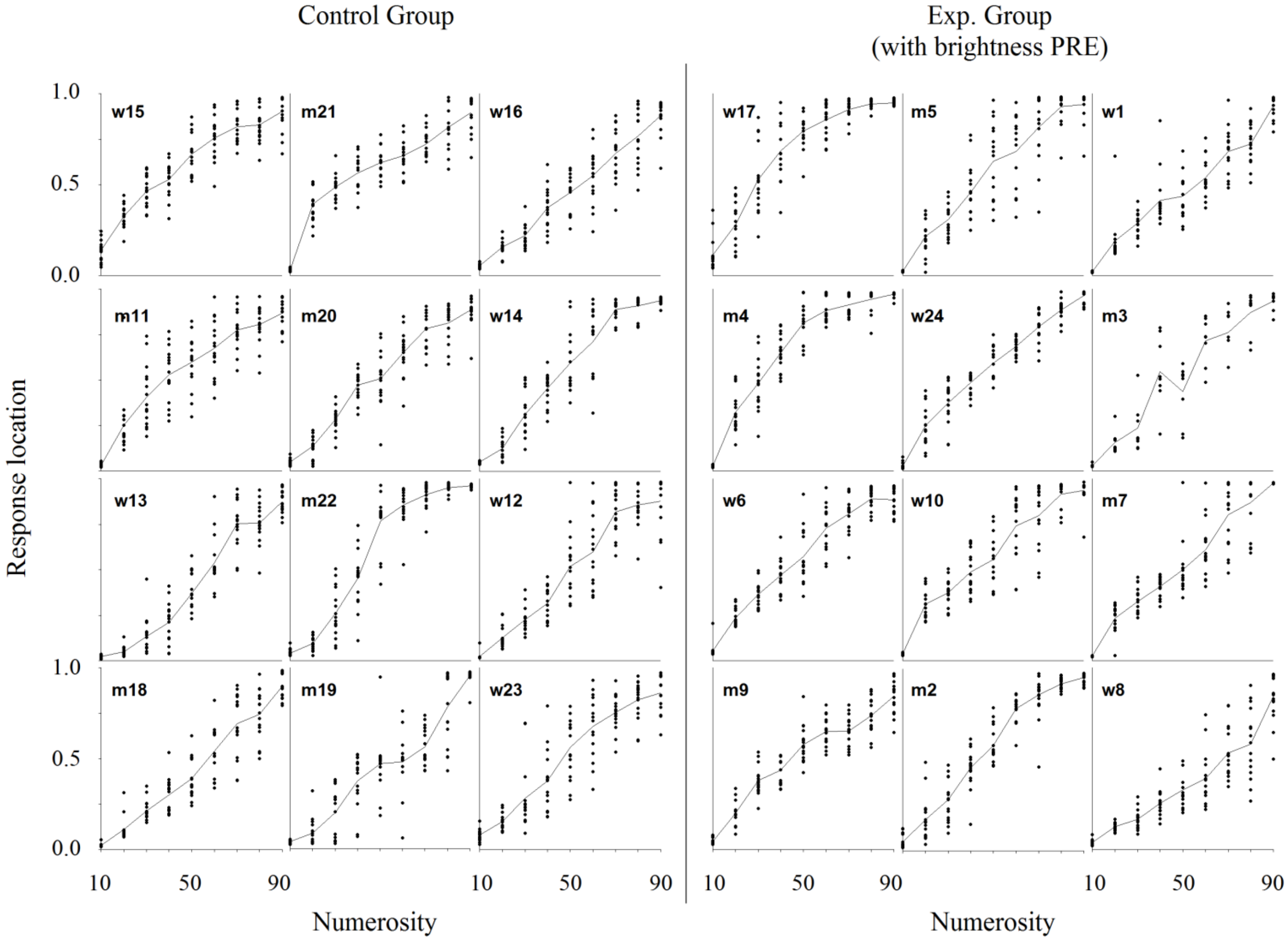
Study 2: Appendix E. Individual trial scatterplots of the number-to-position tasks (Number to position (testing): 2.2 Testing of 2 anchors and 7 intermediate numerosities).

Each dot represents one trial's response, and the line the mean response location.

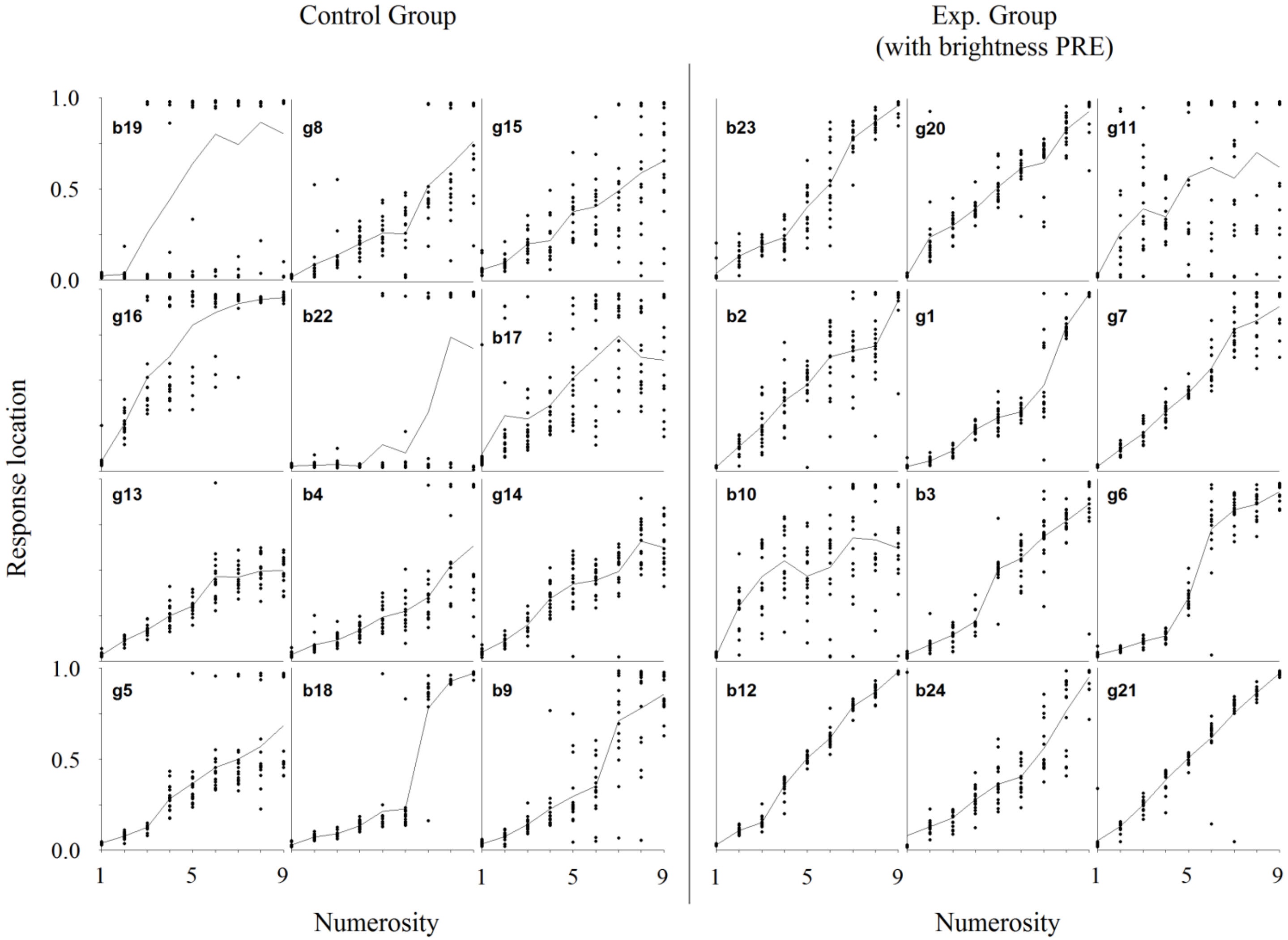
Adults - Dots [1,9]



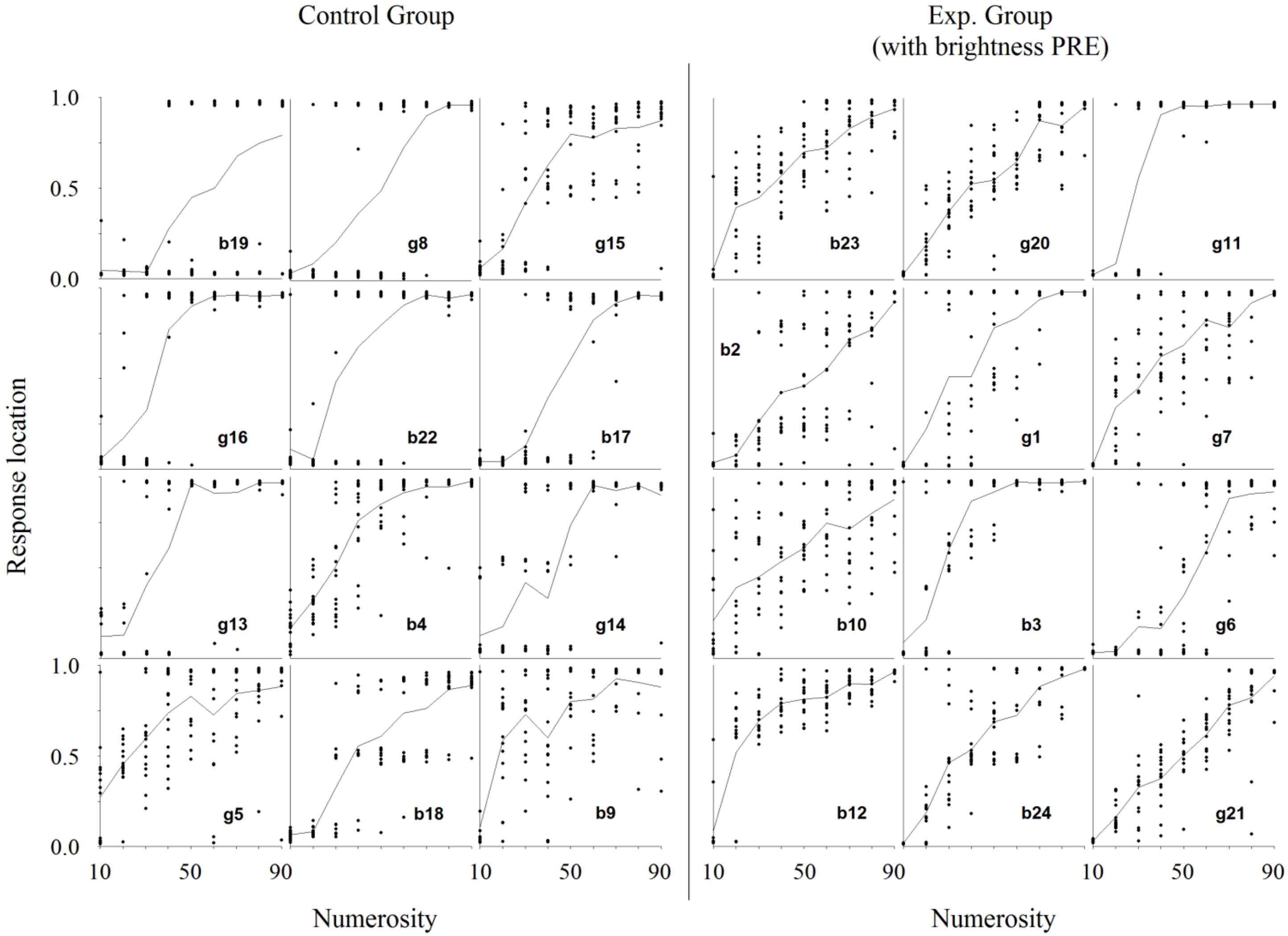
Adults - Dots [10,90]



Children - Dots [1,9]



Children - Dots [10,90]



Study 2: Appendix F. Individual normalized entropy (H) scores.

Entropy was estimated based on the frequency distribution of the response locations, according to the formula: $H = -\frac{\sum_{i=1}^{9}(p_i \times log_2 p_i)}{log_2 9}$.

	Children						**Adults**			
		Age (yrs)	Dots [1,9]	k	Dots [10,90]	k		Age (yrs)	Dots [1,9]	Dots [10,90]
Ctrl										
	b19	4.06	.40	2	.36	2	**w15**	18.13	.96	.97
	g8	4.76	.85	>3	.35	2	**m21**	18.49	.96	.91
	g15	4.95	.90	>3	.74	3	**w16**	18.58	.97	.99
	g16	4.96	.65	>3	.36	2	**m11**	18.98	.98	.98
	b22	5.02	.27	2	.36	2	**m20**	19.57	.99	.98
	b17	5.11	.96	>3	.40	2	**w14**	20.76	.96	.90
	g13	5.28	.81	>3	.42	2	**w13**	21.97	.98	.93
	b4	5.47	.77	>3	.73	>3	**m22**	22.59	.96	.82
	g14	5.50	.90	>3	.47	3	**w12**	22.97	.98	.96
	g5	5.94	.81	>3	.83	>3	**m18**	23.44	.93	.95
	b18	6.07	.60	>3	.63	3	**m19**	24.82	.83	.86
	b9	6.20	.90	>3	.78	>3	**w23**	25.69	.99	.97
	avg	5.28	.74		.54		**avg**	21.33	.96	.93
	SD	0.61	.22		.19		**SD**	2.61	.04	.05
Exp										
	b23	4.12	.91	>3	.93	>3	**w17**	18.20	.96	.83
	g20	4.86	.98	>3	.94	>3	**m5**	18.43	.99	.94
	g11	4.90	.85	>3	.30	2	**w1**	23.09	1.0	.99
	b2	4.93	.97	>3	.84	>3	**m4**	18.66	.98	.87
	g1	4.99	.83	>3	.61	>3	**w24**	18.68	.97	.96
	g7	5.24	.98	>3	.84	>3	**m3**	19.87	.95	.94
	b10	5.27	.96	>3	.93	>3	**w6**	20.42	1.0	.97
	b3	5.54	.92	>3	.47	3	**w10**	21.92	.98	.92
	g6	5.63	.84	>3	.61	>3	**m7**	22.00	.95	.90
	b12	5.85	.94	>3	.72	>3	**m9**	23.27	1.0	.96
	b24	6.18	.93	>3	.81	>3	**m2**	24.92	.93	.92
	g21	6.23	.98	>3	.97	>3	**w8**	25.09	.98	.93
	avg	5.31	.92		.75		**avg**	21.21	.97	.93
	SD	.61	.06		.21		**SD**	2.52	.02	.04

Note. In the k column, '2', '3' and' >3' indicate that the number of clusters that best describe the child's distribution of responses along the bar is 2, 3 or more than three, respectively.

Study 2: Appendix G. “Counting and Verbal Estimation Assessment” recording sheets.

3.2 Counting Objects.

Participant ________________________

Task: serial counting of objects | Range: (1 to 9)

Task: correspondence between marbles and spoken numerals (n ϵ [1, 9]).

I: Experimenter says a number word. Asks: “give me {n} marbles”. ”The child’s response is to give {n} marbles.
II: Experimenter presents the child with a set of {n} marbles. Asks: “how many marbles are there?” The child’s response is to say a number word.
III: Experimenter presents the child with a set of {n} marbles. Asks “give me the same number of marbles as there are here”. The child’s response is to give {n} marbles.

I. Numerals to Marbles	
EXP.	**CHILD**
Numeral asked	Number of marbles given

II. Marbles to Numerals	
EXP.	**CHILD**
{n} marbles presented	Numeral said

III. Marbles to Marbles	
EXP.	**CHILD**
{n} marbles presented	Number of marbles given

Order of conditions	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**
I	1	2	9	3	8	4	7	5	6
II	2	3	1	4	9	5	8	6	7
III	3	4	2	5	1	6	9	7	8
IV	4	5	3	6	2	7	1	8	9
V	5	6	4	7	3	8	2	9	1
VI	6	7	5	8	4	9	3	1	2
VII	7	8	6	9	5	1	4	2	3
VIII	8	9	7	1	6	2	5	3	4
IX	9	1	8	2	7	3	6	4	5

3.3 Verbal Estimation.

Participant _________________________

Range: Smaller (1 to 9) | Larger (10 to 90)

trl	Numerosity of the set of dots	Verbal estimate	Counting (Y/N) 1-by-1 ?	Comments
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				

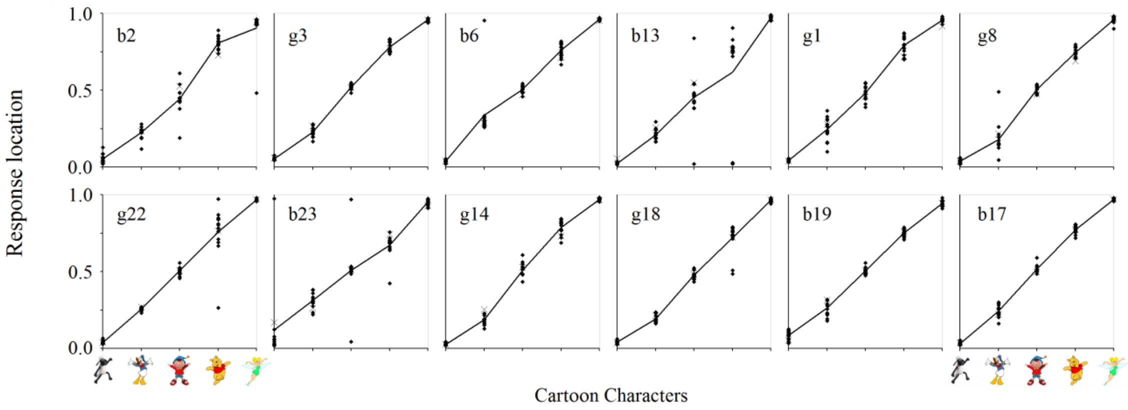
Appendices for Chapter IV: Study 3

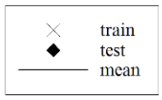
Control of a response continuum by the numerical stimulus continuum: isolating the effect of training responses in different locations of the bar

Study 3: Appendix A. Individual scatterplots of Figures-to-position tasks.

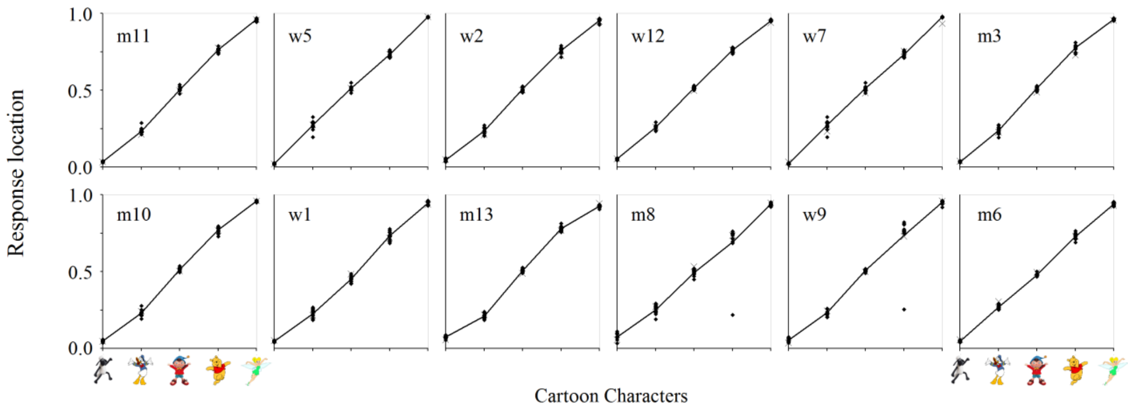
Each data point is one trial's response location. The line depicts mean location. Cross data points represent the mean response locations during the two training sessions, the filled diamonds the responses during the testing session.

Children





Adults



Study 3: Appendix B. Multiple Regression Analyses**.**

Results of OLS multiple regression analyses, showing the unstandardized weights (plus/minus standard error) of the linear (β_{lin}) and logarithmic (β_{log}) regressors, and its corresponding t ratios and p-values (with df = 7), for each stimulus condition.

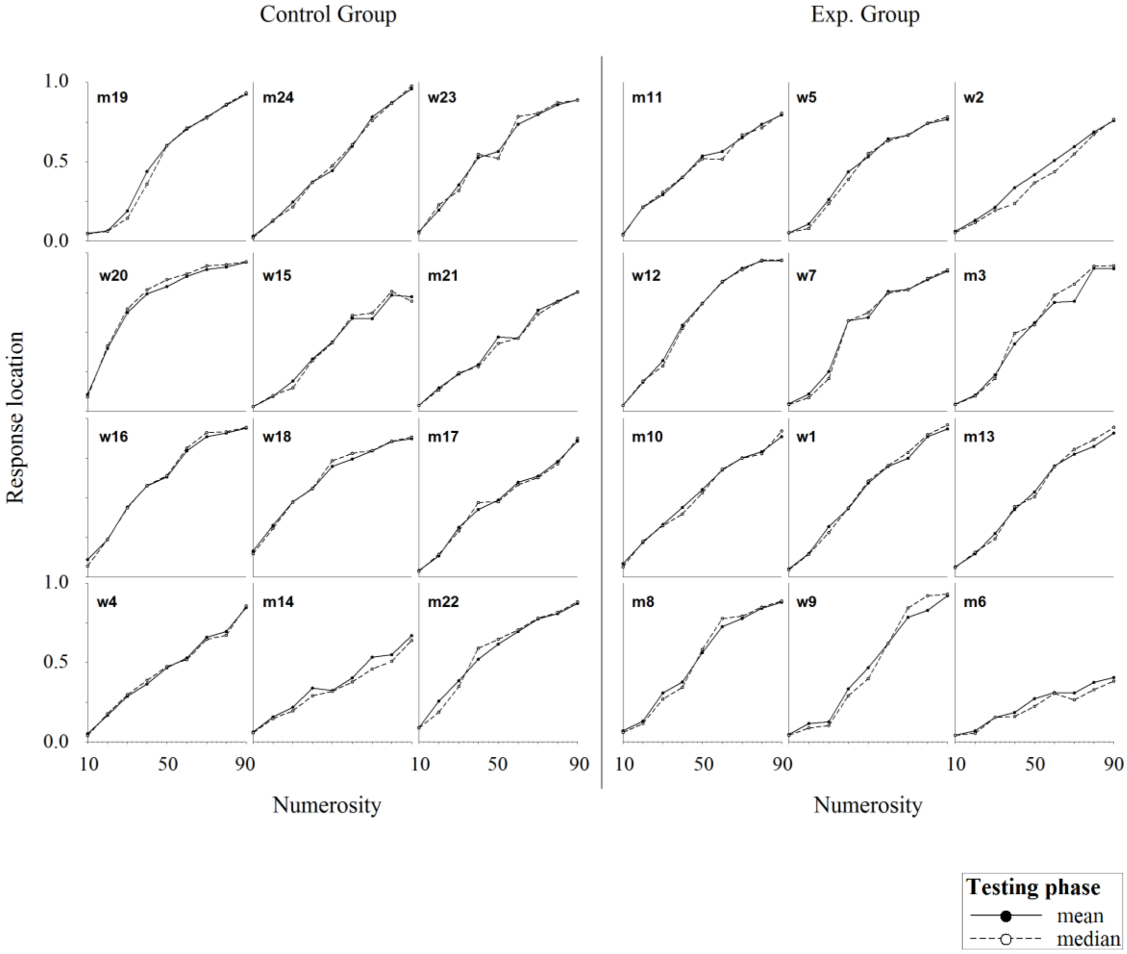
		Dots [10,90]	
		Mean	Median
Adults	Ctrl.	$\beta_{lin} = 0.006 \pm 0.001$ t ratio = 7.349 $p < 0.001$	$\beta_{lin} = 0.006 \pm 0.001$ t ratio = 7.067 $p < 0.001$
		$\beta_{log} = 0.350 \pm 0.109$ t ratio = 4.866 $p = 0.003$	$\beta_{log} = 0.371 \pm 0.127$ t ratio = 3.016 $p = 0.024$
	Exp.	$\beta_{lin} = 0.008 \pm 0.001$ t ratio = 4.900 $p = 0.003$	$\beta_{lin} = 0.009 \pm 0.001$ t ratio = 5.544 $p = 0.001$
		$\beta_{log} = 0.246 \pm 0.065$ t ratio = 1.807 $p = 0.121$	$\beta_{log} = 0.168 \pm 0.082$ t ratio = 0.871 $p = 0.417$
Children	Ctrl.	$\beta_{lin} = 0.011 \pm 0.002$ t ratio = 2.419 $p = 0.052$	$\beta_{lin} = 0.015 \pm 0.003$ t ratio = 1.280 $p = 0.248$
		$\beta_{log} = 0.163 \pm 0.170$ t ratio = 0.401 $p = 0.702$	$\beta_{log} = -0.08 \pm 0.244$ t ratio = 0.039 $p = 0.971$
	Exp.	$\beta_{lin} = 0.005 \pm 0.002$ t ratio = 1.307 $p = 0.239$	$\beta_{lin} = 0.006 \pm 0.003$ t ratio = 0.743 $p = 0.486$
		$\beta_{log} = 0.659 \pm 0.191$ t ratio = 1.896 $p = 0.107$	$\beta_{log} = 0.660 \pm 0.286$ t ratio = 0.884 $p = 0.411$

Study 3: Appendix C. Mean and Median scatterplots of the Number-to-position task.

Graphs are separated according to the experimental group and ordered by participants' age. Black circles connected by the solid line represent the mean, and white circles connected by the dashed line the median response location.

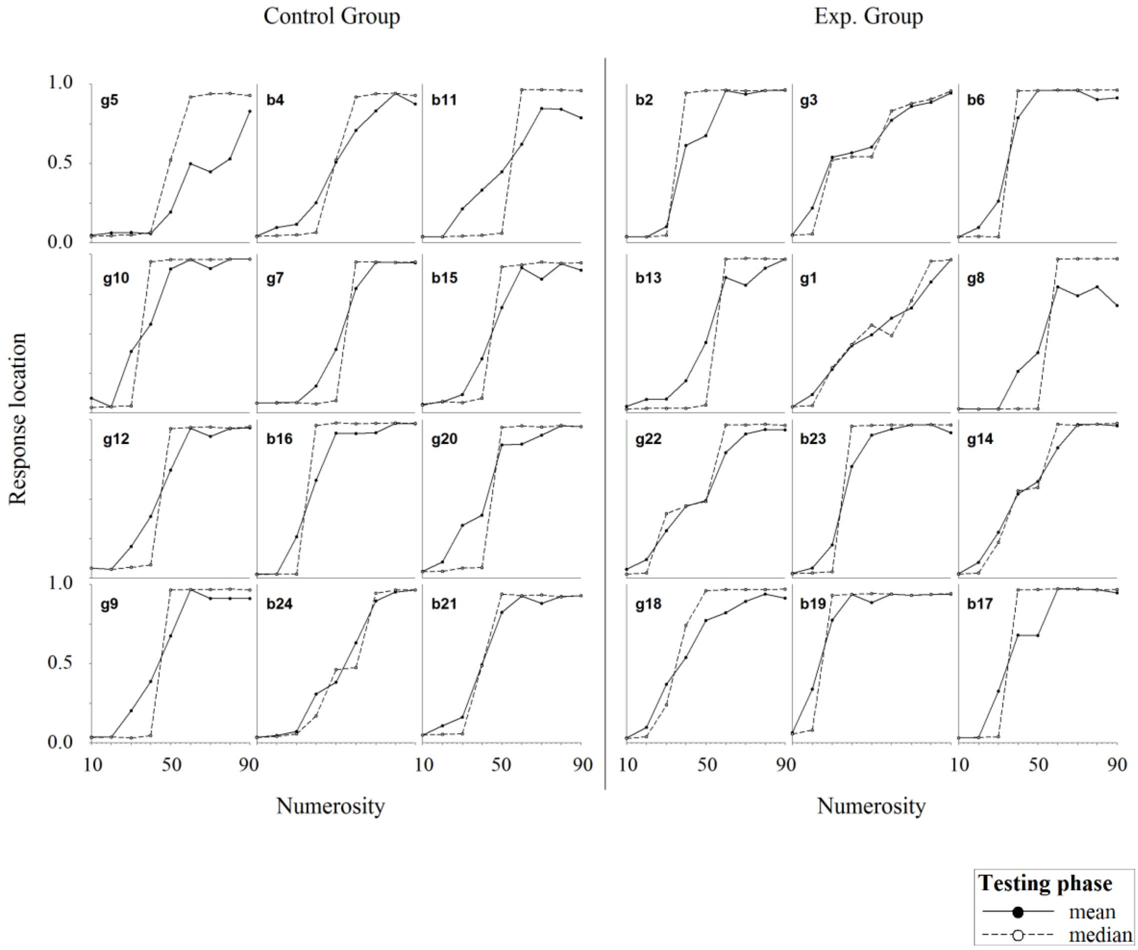
Adults

Dots [10,90]



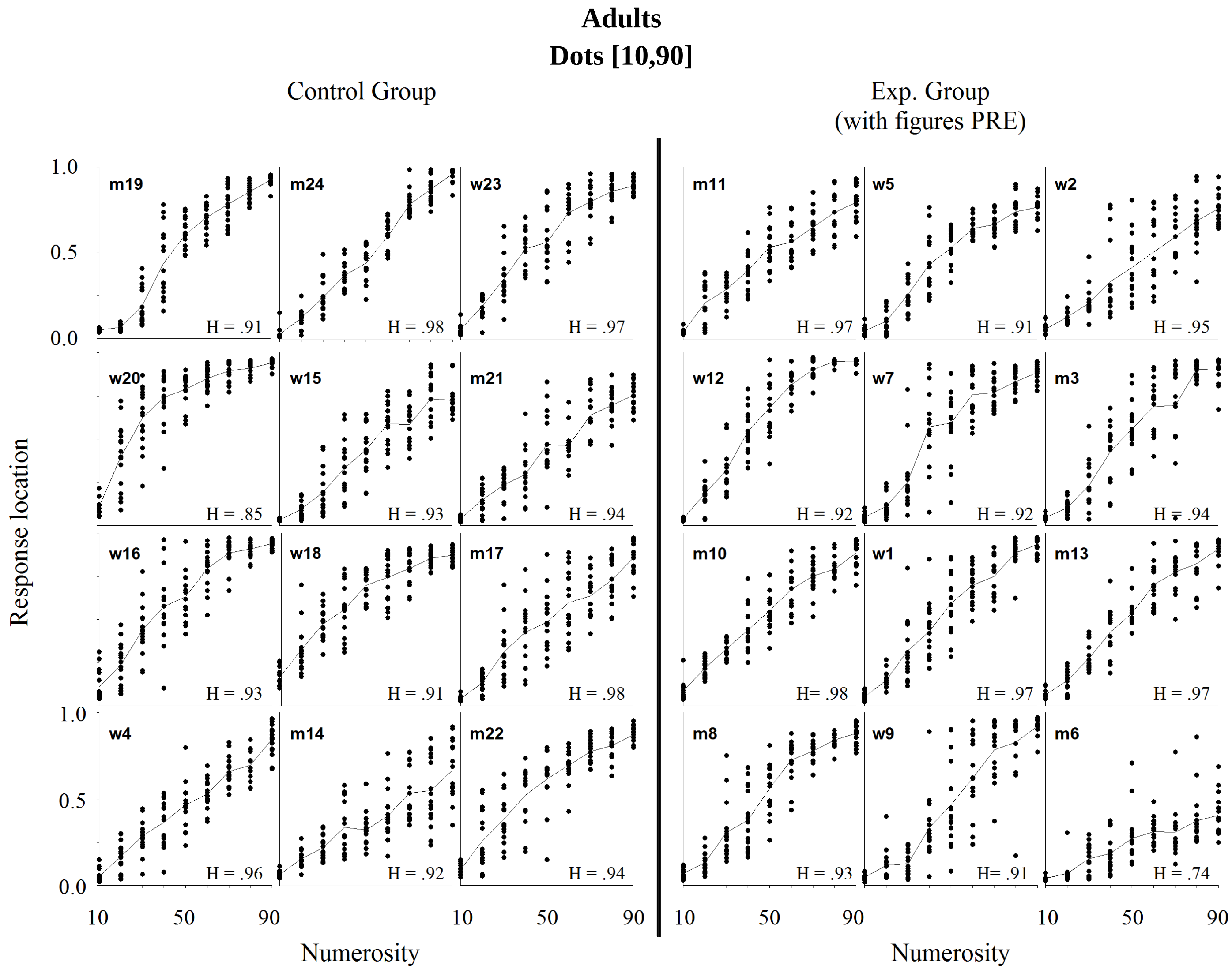
Children

Dots [10,90]

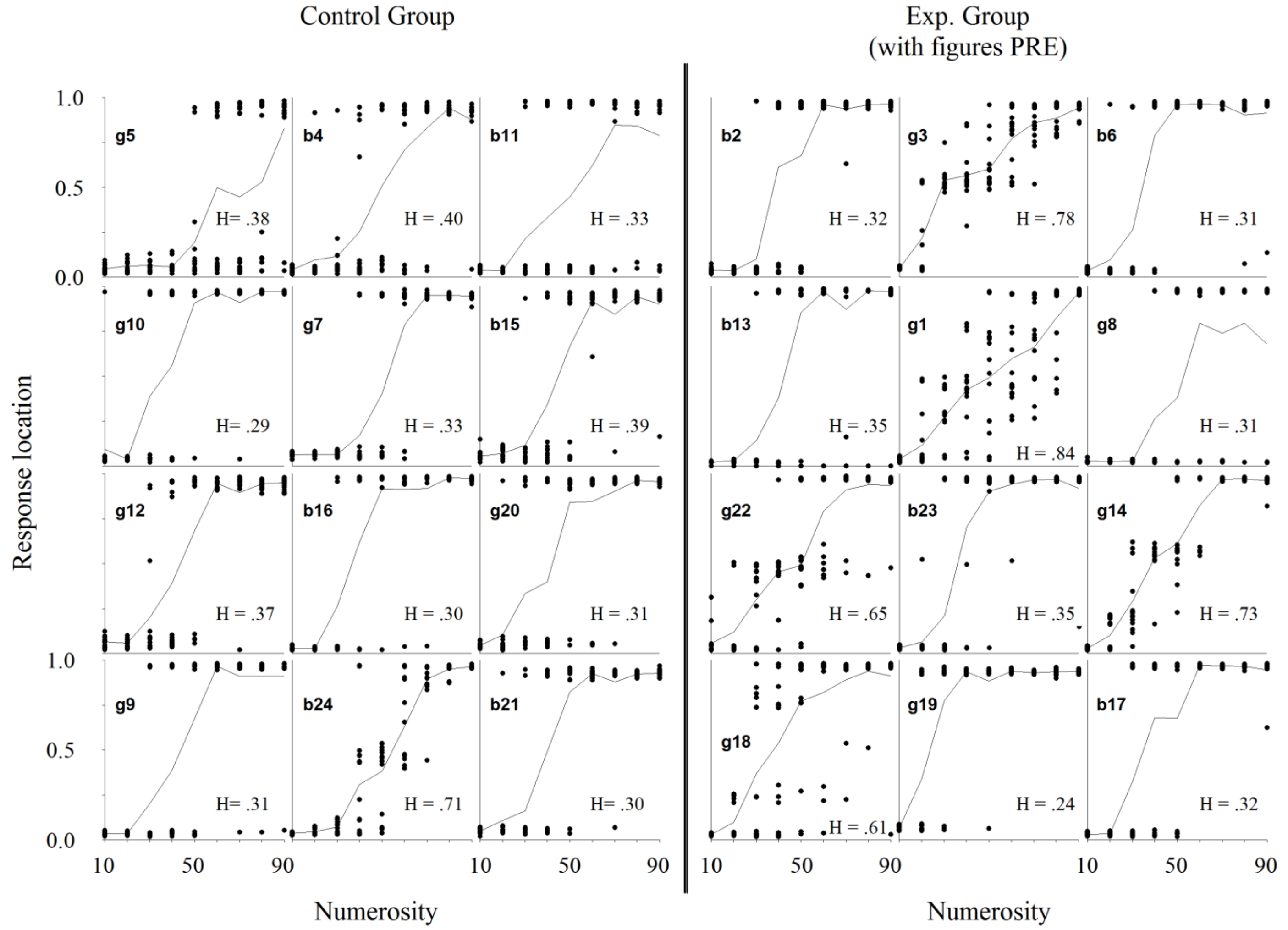


Study 3: Appendix D. Individual trial scatterplots of the Number-to-position task.

Each dot represents one trial's response, and the line the mean response location. Each graph also has the normalized entropy score (H) value.


Adults
Dots [10,90]
Control Group
Exp. Group
(with figures PRE)
Response location
1.0
0.5
0.0
m19
H = .91
m24
H = .98
w23
H = .97
m11
H = .97
w5
H = .91
w2
H = .95
w20
H = .85
w15
H = .93
m21
H = .94
w12
H = .92
w7
H = .92
m3
H = .94
w16
H = .93
w18
H = .91
m17
H = .98
m10
H= .98
w1
H = .97
m13
H = .97
w4
H = .96
m14
H = .92
m22
H = .94
m8
H = .93
w9
H= .91
m6
H = .74
10
50
90
Numerosity
Numerosity

Children
Dots [10,90]



Study 3: Appendix E. Individual normalized entropy scores of the Number-to-position task.

Entropy was estimated based on the frequency distribution of the response locations. For children participants, it is also shown the number of response clusters – k – that best describes the distribution of responses along the bar ('2', '3', or '>3').

	Children			Adults	
		H	k		H
Ctrl.					
	g5	.38	2	**m19**	.91
	b4	.40	2	**m24**	.98
	b11	.33	2	**w23**	.97
	g10	.29	2	**w20**	.85
	g7	.33	2	**w15**	.93
	b15	.39	2	**m21**	.94
	g12	.37	2	**w16**	.93
	b16	.30	2	**w18**	.91
	g20	.31	2	**m17**	.98
	g9	.31	2	**w4**	.96
	b24	.71	3	**m14**	.92
	b21	.30	2	**m22**	.94
	avg	.37		**avg**	.94
	SD	.11		**SD**	.04
Exp.					
	b2	.32	2	**m11**	.97
	g3	.78	3	**w5**	.91
	b6	.31	2	**w2**	.95
	b13	.35	2	**w12**	.92
	g1	.84	>3	**w7**	.92
	g8	.31	2	**m3**	.94
	g22	.65	3	**m10**	.98
	b23	.35	2	**w1**	.97
	g14	.73	3	**m13**	.97
	g18	.61	>3	**m8**	.93
	g19	.24	2	**w9**	.91
	b17	.32	2	**m6**	.74
	avg	.48		**avg**	.93
	SD	.22		**SD**	.06

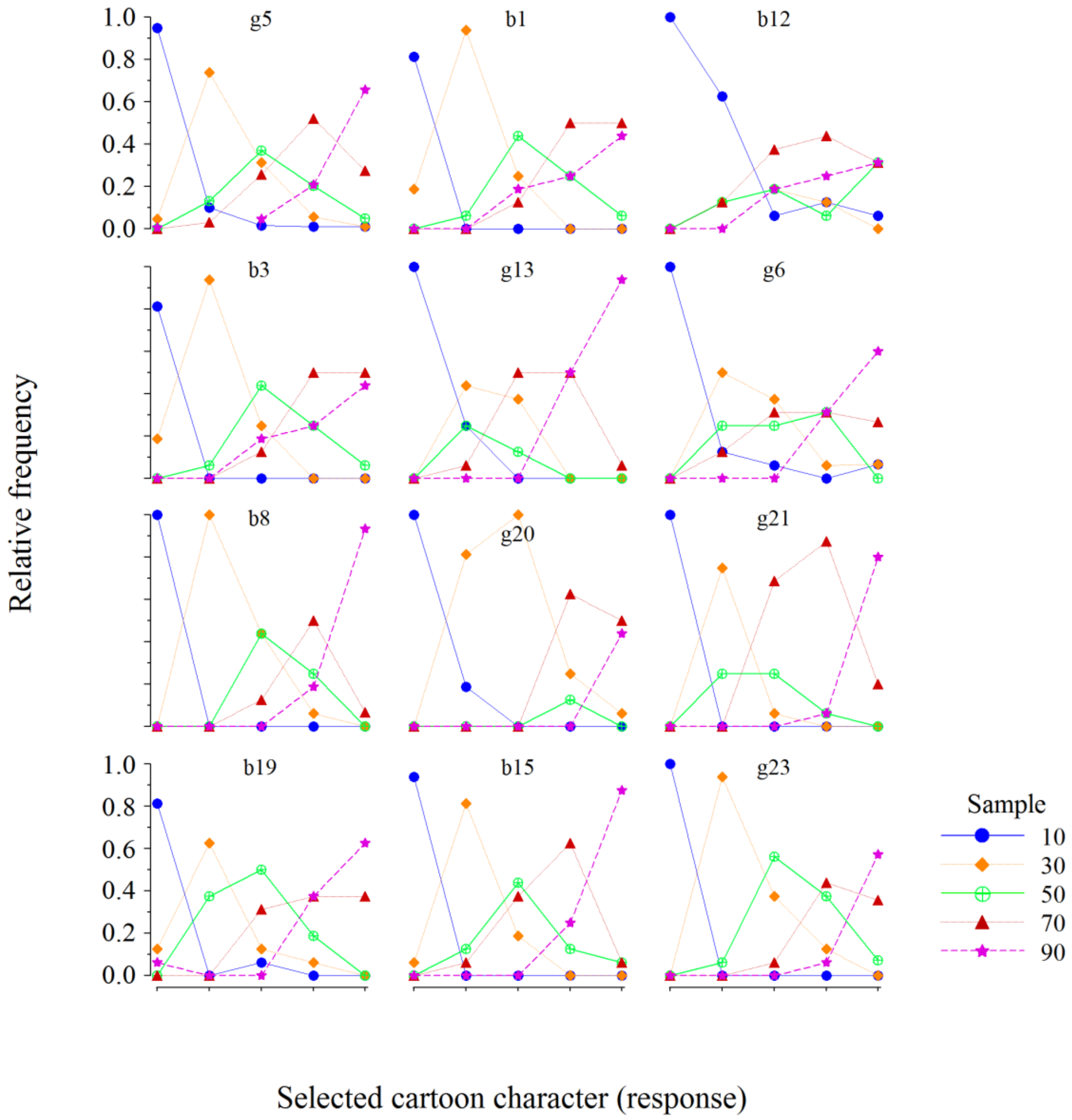
Appendices for Chapter V: Study 4

Control of a response continuum by the numerical stimulus continuum: isolating the effects of a perceptual training on Number-to-Position performance

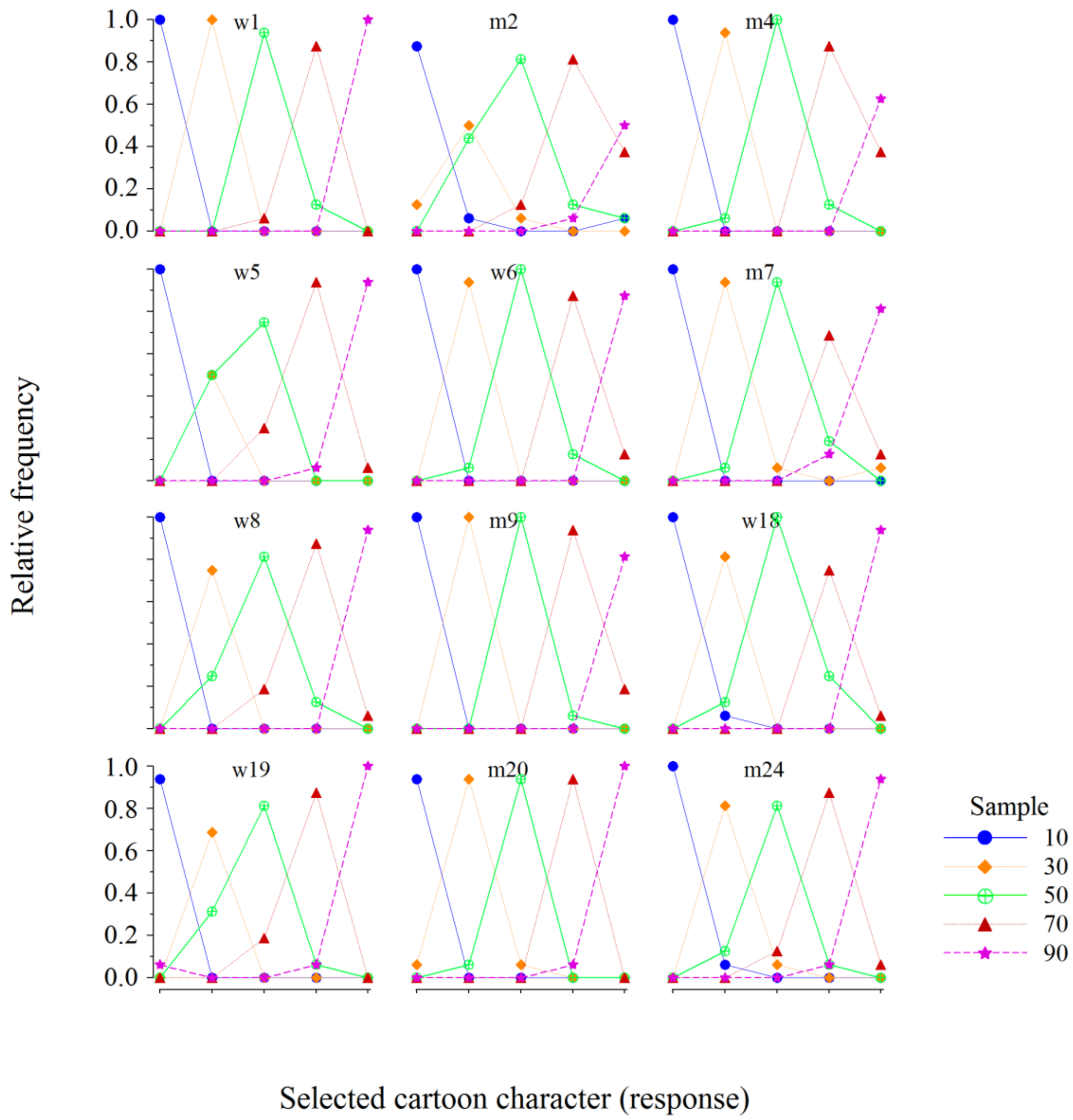
Study 4: Appendix A. Individual performance during the "Perceptual Training" (test session).

Each colored series represents how often each of the five presented images was selected, after the presentation of a specific sample (numerosity).
The correct Number-Image assignments were: '10-Shaun', '30-Donald', '50-Noddy', '70-Winnie', and '90-Tinkerbell'.

Children



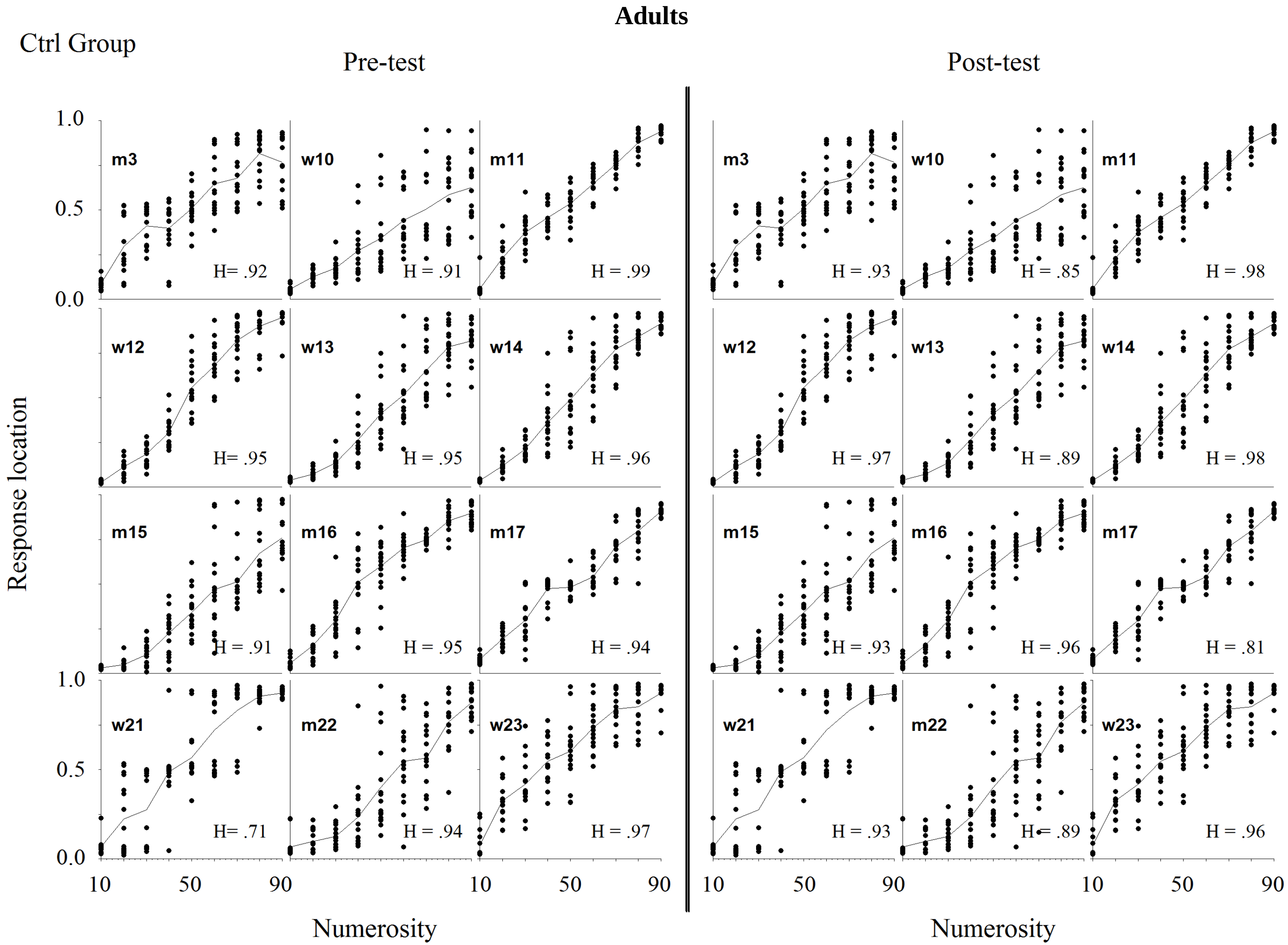
Adults

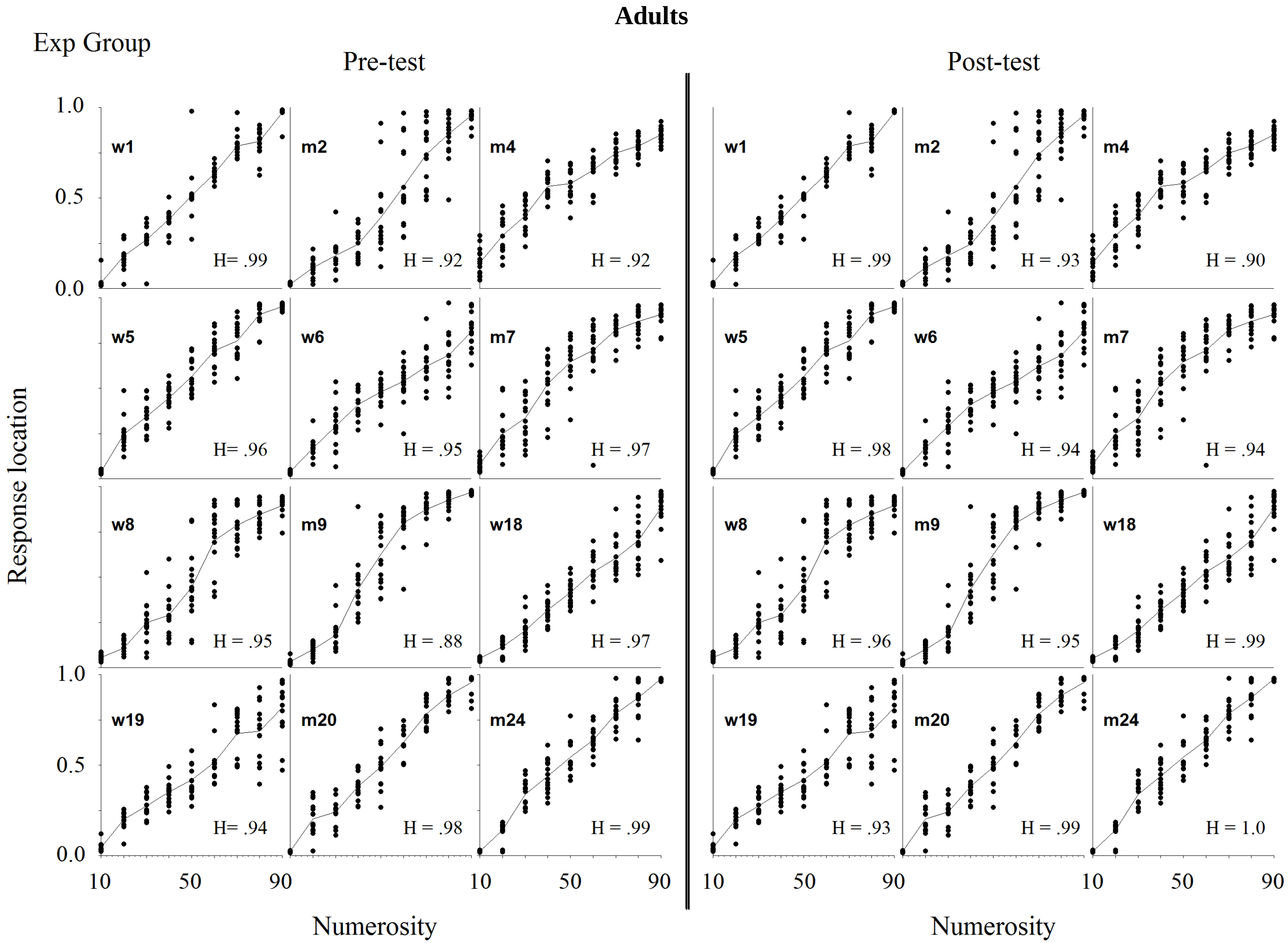

1.0
0.8
0.6
0.4
0.2
0.0
w1
m2
m4
w5
w6
m7
w8
m9
w18
w19
m20
m24
Relative frequency
Selected cartoon character (response)
Sample
10
30
50
70
90

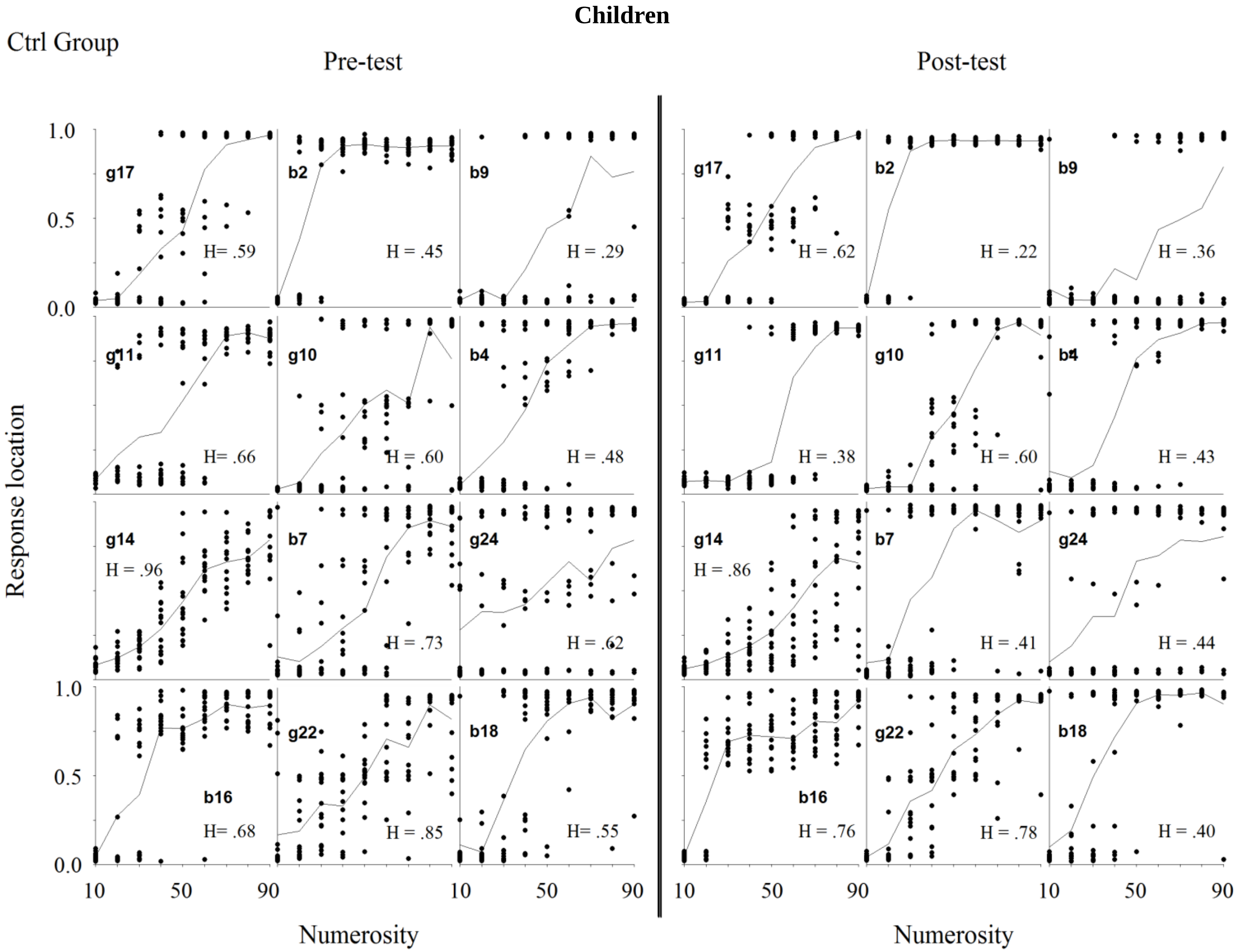
Study 4: Appendix B. Individual trial scatterplots of the Number-to-position tasks (Pre- and Post-test sessions).

Each dot represents one trial's response, and the line the mean response location. Each graph also has the normalized entropy score (H) value, calculated according to the formula: $H = -\frac{\sum_{i=1}^{9}(p_i \times log_2 p_i)}{log_2 9}$.

.

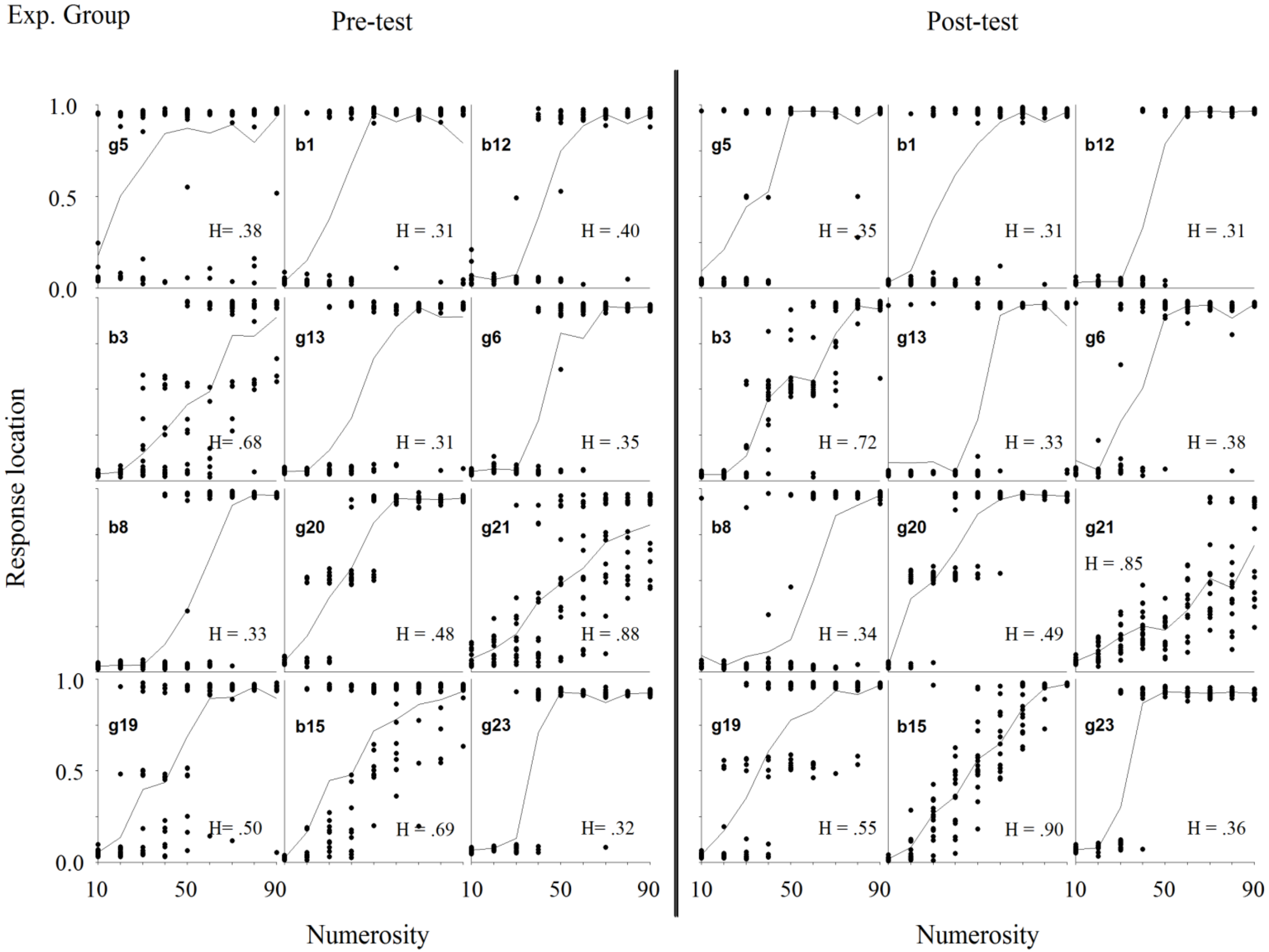

Adults
Ctrl Group
Pre-test
Post-test
Response location
1.0
0.5
0.0
m3
H= .92
w10
H = .91
m11
H = .99
m3
H = .93
w10
H = .85
m11
H = .98
w12
H= .95
w13
H = .95
w14
H = .96
w12
H = .97
w13
H = .89
w14
H = .98
m15
H = .91
m16
H = .95
m17
H = .94
m15
H = .93
m16
H = .96
m17
H = .81
w21
H= .71
m22
H = .94
w23
H = .97
w21
H = .93
m22
H = .89
w23
H = .96
10
50
90
Numerosity
Numerosity


Adults
Exp Group
Pre-test
Post-test
Response location
Numerosity
1.0
0.5
0.0
10
50
90
w1
m2
m4
w5
w6
m7
w8
m9
w18
w19
m20
m24
H= .99
H = .92
H = .92
H= .96
H = .95
H = .97
H = .95
H = .88
H = .97
H= .94
H = .98
H = .99
H = .99
H = .93
H = .90
H = .98
H = .94
H = .94
H = .96
H = .95
H = .99
H = .93
H = .99
H = 1.0



Children
Ctrl Group
Pre-test
Post-test
Response location
Numerosity
Numerosity
g17
H= .59
b2
H = .45
b9
H = .29
g17
H = .62
b2
H = .22
b9
H = .36
g11
H= .66
g10
H = .60
b4
H = .48
g11
H = .38
g10
H = .60
b4
H = .43
g14
H = .96
b7
H = .73
g24
H = .62
g14
H = .86
b7
H = .41
g24
H = .44
b16
H= .68
g22
H = .85
b18
H= .55
b16
H = .76
g22
H = .78
b18
H = .40

Children


Exp. Group
Pre-test
Post-test
Response location
1.0
0.5
0.0
g5
b1
b12
b3
g13
g6
b8
g20
g21
g19
b15
g23
H= .38
H = .31
H = .40
H = .68
H = .31
H = .35
H = .33
H = .48
H = .88
H= .50
H = .69
H= .32
H = .35
H = .31
H = .31
H = .72
H = .33
H = .38
H = .34
H = .49
H = .85
H = .55
H = .90
H = .36
10
50
90
Numerosity
Numerosity

Appendix E: Temporal and Numerical Bisection

Title

The effect of practice on behavioral sensitivity to temporal and numerical stimuli using the Bisection task

Note: this work was presented during the Training School "Timing and Time Perception: Procedures, Measures, & Applications", held between February the 4th and 8th, 2013, at the Ionian University in Corfu (Greece). E. Fernandes applied for and received the COST Action TD0904 grant from the COST scientific program on *Time in Mental activity: theoretical, behavioral, bioimaging and clinical perspectives* (*TIMELY*), that covered the travel and housing expenses to attend this Training school.

Introduction

This study aims to investigate how sensitivity to temporal and numerical stimuli is improved by practice and, additionally, whether a mere exposure to the test stimuli or a perceptual training produced differentiated degrees of amelioration. Namely, we analyze how adults' performance in bisection tasks changes following a discriminative training or simply the presentation of the relevant stimuli.

The experiment will consist of three phases: (1) Pre-test – 1st Bisection task, (2) Treatment, and (3) Post-Test – 2nd Bisection task. During Pre-test, adults will receive a standard bisection procedure, in which after learning the correct assignments between two anchor stimuli and a response ("Short" or "Long"), they will be tested with intermediate values. Participants will then be divided in three groups during the Treatment phase. Participants in the (2.1) Exposure group will be presented the stimuli used in the bisection phase. Those in the (2.2) Training group will be presented the stimuli but at each trial, rather than classifying the stimuli as "Short"/"Long", they will be trained to select an appropriate figure amongst five presented (flower, apple, guitar, glove and tweezers). Thus, they will experience categorical perception training (Goldstone & Hendrickson, 2009) within the numerical range of the bisection task. The (2.3) Control group will not entail a treatment, merely a temporal delay that equates the other two groups' experimental duration. In the final Post-Test phase, we will repeat the events of the first bisection session.

Individual performance in the Post-Test phase (2nd bisection session) will be contrasted to the baseline, in terms of sensitivity (Weber Ratio) and bias (Point of Subjective Equality). We will also analyze inter-subject differences between the three treatment groups. Finally, due to the growing evidence on the similarity between the mechanisms of time and numerical discrimination (Fernandes & Church, 1982; Meck & Church, 1983; Roberts, 1995, 2006; Roitman, Brannon, Andrews, & Platt, 2007; Brannon, Suanda, & Libertus, 2007; Droit-Volet, Clément, & Fayol, 2008; Allman, Pelphrey, & Meck, 2012), the experiment will be conducted with temporal and numerical stimuli.

To our knowledge, no study has focused in the effects of perceptual learning (Gibson, 1969) on humans' nonverbal numerical discrimination. But ever since the 1950's there has been evidence that categorizing (or identifying) stimuli can increase perceptual sensitivity in position, area, length, hue, saturation and brightness discrimination (Eriksen & Hake, 1955a, 1955b; Baird et al., 1970; Hake & Garner, 1951; Goldstone, 1994b; Burns & Shepp, 1988; Ozgen & Davies, 2002; Roberson, Davidoff, & Davies, 2005; Winamer et al., 2007).

In the current study, participants in the Training group will receive a categorical perception training protocol (Goldstone, & Hendrickson, 2009). Our participants will have to make discriminations along the numerical range for which sensitivity will be assessed in a bisection task. In our training protocol, the tag to express each concept is a specific image, rather than a verbal tag as would be the case if we asked subjects to verbally estimate the number of dots. Our hypothesis is that, compared to a Control group, the psychometric curves from the participants who receive this differentiation or “tagging” would present higher sensitivity to number or time.

We are less certain about what to expect of the Exposition treatment. On the one hand, some evidence shows that mere exposure is not sufficient to increase expertise (James, 1890). On the other hand, overt discrimination training is not always necessary for an increase in perceptual sensitivity. Both “preexposure effect” (Hall, 1991, 2001) and “predifferentiation” (Gibson (1991) perceptual studies have demonstrated that the presentation of the tested materials, without feedback, may increase responses’ sensitivity to those stimuli.

Methods

Participants. The experiment studied sixty Portuguese adults (30 women), aged between 18.3 and 53.3 years-old (M = 35.4, SD = 11.3) Informed consent was given by all participants. Participants were voluntaries who agreed to participate, without gaining any external reward.

We followed a convenience sampling method. Due to time constraints of the experimenter's schedule, who in those days was also collecting data in a kindergarten during day-work hours, subjects were selected because of their accessibility and proximity to the experimenter. Also, though most participants were directly contacted by the experimenter, there was also a "snowball sampling", for she often asked current participants to refer or identify other possible subjects among their relatives or friends. Data collection occurred mostly during after-work hours (i.e., after 6 p.m.), either in a separate room made available at the Sports Center of Martim (a village in the North of Portugal) or in the residence of a local family.

Numerical stimuli. The same numerical stimuli that were created for the Dots[10,90] condition in the mapping experiments were used on the current bisection study. Briefly, the numerical samples were the presentation of green circles, centered on screen, which contained a variable number of red dots. The samples could be sets of 10, 20, 30, 40, 50, 60, 70, 80, or 90 dots. Please refer to the Methods section of Study 2 for the complete description of how we created the numerical arrays.

Temporal stimuli. During the temporal bisection tasks, during the sample moment a red square, side 300 pixel, was presented centered on screen. The duration that this square was presented constituted the temporal sample. The square could be on screen during 400, 800, 1200, 1600, 2000, 2400, 2800, 3200, or 3600 milliseconds (ms). Note that the ratios between the smallest and largest duration (400 vs. 3600) and between the smallest and largest numerosity (10 vs. 90) are both of 1:9.

Procedure. Participants were seated in front of a touchscreen laptop (HP Pavilion tx2000 Notebook PC, screen size 12.1", screen resolution 1024 x 768, refresh rate 60 Hz), in a quiet room. Experimental programs written using the E-Prime software (Version

2.0, Pittsburgh: Psychology Software Tools Inc.) controlled all session events and recorded participants' responses during the bisection tasks. A program written in Visual-Basic language was also implemented during the treatment phases introduced in-between the bisection tasks in the "Exposure" and "Training" groups' procedure. When the task was numerical, two pink square stickers, one with character 'P' printed in black, the other with 'M', were fixed upon the keyboard's keys 's' and 'l'. When the task was temporal, the printed characters were 'C' and 'L'. The assignment of a sticker to each key was counterbalanced across subjects. A yellow rectangular sticker was placed upon the spacebar of the keyboard. The experimenter remained in the room with the participant, but seated herself about 1 m behind the participant to keep out of his sight and prevent response bias. A separate monitor, positioned behind the participant and facing the experimenter, was connected to the laptop and displayed the experimental events.

A participant either completed the experiment with only numerical stimuli (Number condition) or solely with durations (Time condition). Thus, half the participants were assigned to each stimulus dimension. Within each condition, participants were further divided into three experimental groups. Participants were distributed among the experimental groups in a way that made it as close as possible for the groups to be matched on the number of males and the participants' age. The experimental design is illustrated in Figure 1.

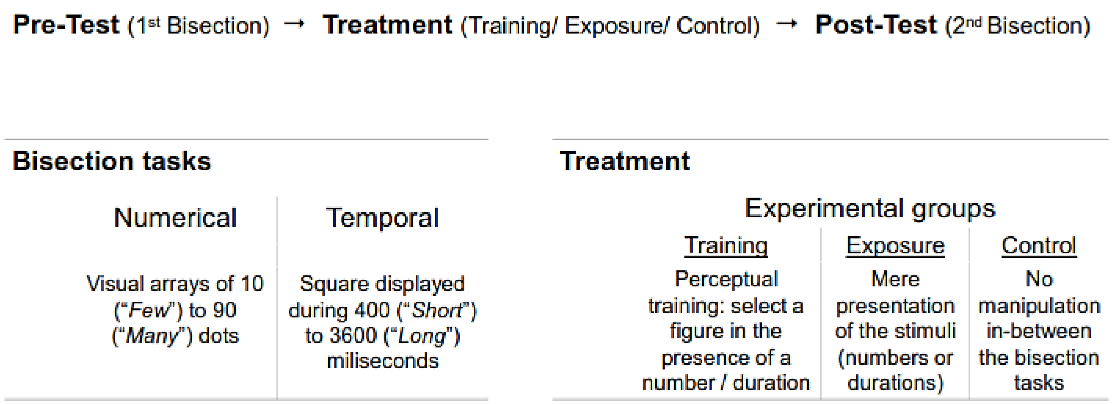


Figure 1. Experimental design. Participants first completed one bisection session (Pre-test). Next, depending on their experimental group, they received one of three possible treatments. Finally, they repeated the bisection task (Post-Test).

For simplicity's sake, we will describe the procedural events for a participant assigned to the Numerical condition. The procedure follows an ABA design or, in other words, is a case of a pretest-posttest control-group research design, with two experimental groups. Participants from both groups were pre tested on their number bisection performance with sets of 10 to 90 dots and later post tested after the experimental treatment condition was administered to the groups.

1. Bisection task (Pre-test). In the very first experimental session, participants seated themselves in front of the computer and read the following instruction:

"*Thank you for participating in our study. You will be presented with the task of evaluating numerosity quantities, presented in the form of sets of dots.*

You must never count the dots. If you do so, you will be distorting the results. Do not count!

(press the yellow bar to continue)

You will be presented with 2 numerosity quantities: one quantity with few points (FEW) and the other with many points (MANY). Your task will be to report if the quantity observed is more similar to the quantity with few points (FEW) or to the quantity with many points (MANY)."

The participants then read another set of instructions in which they were told that they would be presented with the two anchor quantities. Following the indication to the FEW anchor, four exemplars of the FEW quantity ('10') were successively displayed; and the same happened for the MANY quantity ('90').

Afterwards, the training phase began with the following instruction appearing:

"*You will now learn how to respond to each of the two numerosity quantities.*

If it is the quantity "FEW", you must press the key P for POUCO (translation: Few).

If it is the quantity "MANY", press key M for MUITO (translation: Many).

Following each response, you receive a feedback indicating whether your response is correct or no."

On this training phase participants learned the mapping between each anchor numerosity and each response: following the FEW and MANY quantities, press 'P' or 'M' key, respectively. Each trial began with the screen white and an inter-trial interval (ITI), whose duration was a random number between 500 and 1000 ms. Afterwards, the word "ready?" (translation for "pronto?") appeared on the screen, and the participants had to press the yellow bar which, after a delay of 75 ms, resulted in the presentation of the numerical stimuli (sets of dots) during 1500 ms. Then the screen turned white and the subject responded to one of the "P" or "M" sticker-keys. Each response was followed by the feedback stimuli, according to response accuracy. Feedback stimuli consisted of two circular images, 390 pixel diameter, centered on the screen and displayed during 1000 ms. The feedback stimulus that followed correct responses was a yellow smiling face over a white background, while incorrect responses were followed by a frowning face. The training phase ended when the participant had produced 6 correct responses during the previous 8 trials.

The test phase comprised ten 9-trial blocks. Within each block, the trials were randomly ordered. A block of test trials was composed by one exemplar of each numerosity: the two trained anchors and seven arithmetically spaced intermediate numerosities (i.e., sets of 10, 20, 30, 40, 50, 60, 70, 80 and 90 dots). Trials' events during testing were similar to training ones', with the two exceptions that (i) novel numerical values could be presented, and (ii) no response was ever followed by feedback, so that the participant's response immediately terminated the trial. Accordingly, before starting the test phase participants read the instruction:
"*Next, you will be presented with a variety of quantities, but your task is the same. You must judge whether the quantity is more similar to the 'FEW' quantity or more similar to the 'MANY' quantity. At times, you will receive feedback to the quantities FEW and MANY. Don't forget to respond as fast as possible and do not count*!".

2. Treatment. About one hour after completing the bisection task, participants in the "Exposure" and "Training" groups returned to the room and sat at the computer for a new experimental phase.

2.1) Training Group. In this experimental manipulation, five of the numerical stimuli that had been presented in the bisection task (10, 30, 50, 70, and 90) now became discriminative stimuli (S^D) to select one among five figures of objects (apple, flower, guitar, gloves, and pliers). At the beginning of each trial, the computer screen was uniformly colored in light blue (RGB color (147; 204; 234)) and after an ITI of 500 ms, a star image (diameter about 3 cm) appeared in a random location. A finger touch to this image started the presentation of the numerical sample (i.e., a set of dots). The numerical set was presented horizontally centered and about 0.1 cm below the upmost part of the screen. After the sample onset (250 ms later), five square picture boxes, with side 2.5 cm, were shown horizontally aligned at the lowest part of the screen. Each picture box contained an image so that figures of five objects (apple, flower, guitar, gloves, and pliers) were displayed below the numerical sample. The distribution of the images among the five picture boxes was randomly distributed across trials. A touch to a picture box was signaled by the appearance of a yellow inverted triangle above it, similar to an arrow pointing to the selected figure. Whether or not the selected figure had been the correct one was signaled to experimenter in the other screen, hidden from the participant. The experimenter then provided the accurate feedback to the participant's choice: in the case of a correct response she complimented the participant and, in the case of an incorrect choice, informed him/her that the answer was incorrect and that they would repeat the trial. If the response had been correct, a new trial began with the ITI. In case of an incorrect response, a correction procedure was in effect. This correction procedure was just as described for the Number-to-position experiments, namely, a two-step loop, where in the first repetition the experimenter taught the correct answer and in the second repetition, the participant had to respond on his own. This training progressed until the participant had reached at least five consecutive correct responses for each of the five numerical samples. At the end of this session the researcher wrote down the total number of novel trials (i.e., not counting the repeated trials due to the correction procedure) so it would be used as the maximum number of trials for the participant in the 'Exposure' group. In other words, we attempted to yoke the number of stimuli presentations between the 'Training' and 'Exposure' groups.

2.2) Exposure Group. In this treatment manipulation, participants came to the room and sat at the computer. Again, the trial began with the participant touching the star image, and then he saw the numerical sample and selected one of the five picture boxes. However, and differently from the 'Training' manipulation group, all picture boxes contained the image of the to-start star. A touch to any of the picture boxes concluded the trial and participants went through as many trials as their yoked training-counterpart had gone. This way, the participants from the 'Exposure' group received a similar number of stimulus presentations (numerical samples) to participants in the 'Training' group, but unlike the later, they were not trained to "use" these presentations in a stimulus control learning situation (i.e., assign each numerosity to a specific object).

2.3) Control Group. Participants in the 'Control' group received no treatment phase. Merely, they experienced the same time delay as did the participants from the other experimental groups, until they were re-tested in the second bisection task.

3. Bisection task (Post-test). After an hour had elapsed since completion of the treatment condition, participants in the 'Training' and 'Exposure' groups returned to the experimental room. They received the same experimental procedure as was described for the first bisection task. In other words, they repeated the bisection session. As for participants from the 'Control' group, they performed this second bisection task at about the same distance time that the other participants had had after their first task's end.

As for the other half of the participants, who were tested with temporal stimuli, the procedure was the same with the exceptions that the instructions and sticker-keys were adapted to present the smallest and largest durations with the "*Short*" ("*Curto*", in Portuguese) and "*Long*" ("*Longo*") verbal tags. The sample, as described above, consisted in the presentation of a red square centered on screen during 400 to 3600 milliseconds.

Results and Discussion

The main objective of the current work was to test for an effect of two experimental manipulations with numerical and temporal stimuli – exposure or perceptual training – in a subsequent discrimination test. Additionally, given the considerable scarce number of bisection studies, in comparison with other discrimination procedures, we were also interested in inspecting psychophysical discriminations, per se.

We will start by presenting the results of the implementation of the 'Training' treatment group. Next we will address, both at the group and individual level analyses, the psychometric curves obtained in the bisection sessions.

Performance in the Treatment manipulations

All participants who received the 'Training' treatment successfully reached the learning criterion: to emit five consecutive correct responses for each of the five samples (refer to Figures 3 and 4 for the complete individual results). The group results are summarized in Figure 2, where the colored series represent how often each of the five presented images was selected by the participants, after a specific sample (numerosity or duration) was presented to them. The sample codes in Figure 2's legend are ordered by increasing magnitude, so that sample ‘a’ refers to a set of 10 dots in the Number condition, or a duration of 400 ms in the Time condition, whereas the sample ‘e’ refers to the largest sample, a set of 90 dots or a duration of 3600 ms. Although the correct Sample-Image assignments were counterbalanced across subjects, for the sake of a simpler illustration we depict the results as if all subjects were trained with the following assignments: a-Apple, b-Gloves, c-Guitar, d-Pliers, and e-Flower.

In Figure 2, each colored series peaks at the correct image, at frequency values above .81. The few errors that occurred were mostly responses at the immediate vicinity of the correct stimulus. In other words, when participants erred they were selecting the images that were associated with the sample value immediately above or under the correct one. Yet another feature of this phase was that performance in terms of accuracy was very similar between the two stimulus dimensions. In fact, the mean percentage of correct responses was 89% in both Number and Time ‘Training’ conditions. Such similarity in regards to accuracy was also expressed in the number of trials required to reach the learning criterion, an average of 61.7 (SD = 23.0, min. = 40, max. = 103) in the Number dimension, and of 67.3 (SD = 32.8, min. = 26, max. = 136) in the Time dimension (independent samples t-test, $t(1,18) = -5.5$, $p = .664$).

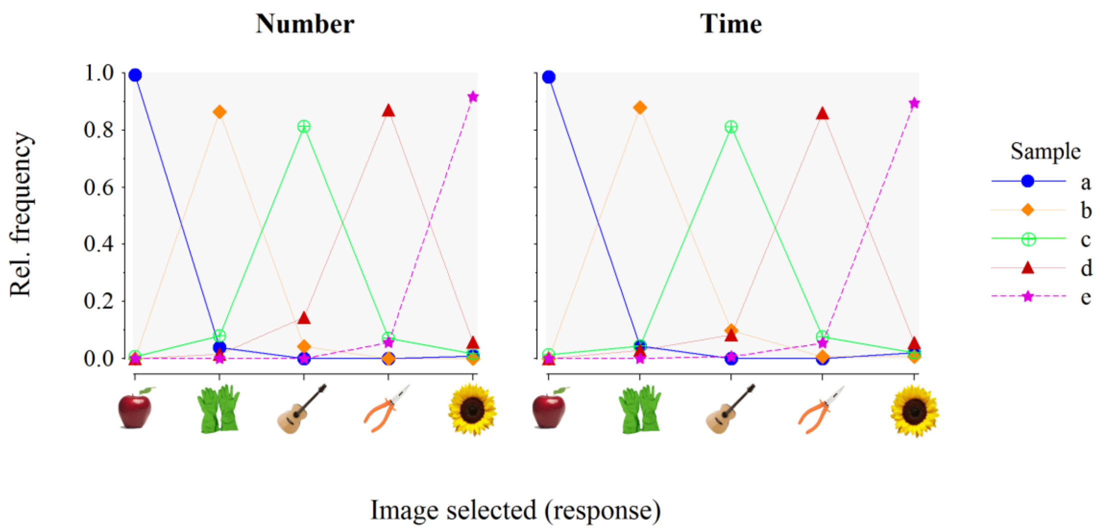


Figure 2. Relative frequency of the images selected during the ‘Training’ treatment conditions. The colored data point series specify following which sample (numerosity or duration) were the images selected, given that the correct Sample-Image assignments were a-Apple, b-Gloves, c-Guitar, d-Pliers, and e-Flower. Sample codes in the legend (letters ‘a’ to ‘e’) are ordered by increasing magnitude (i.e., Number: ‘a’ = 10 dots, ‘e’ = 90 dots; Time: ‘a’ = 400 ms, ‘e’ = 3600 ms).

Number

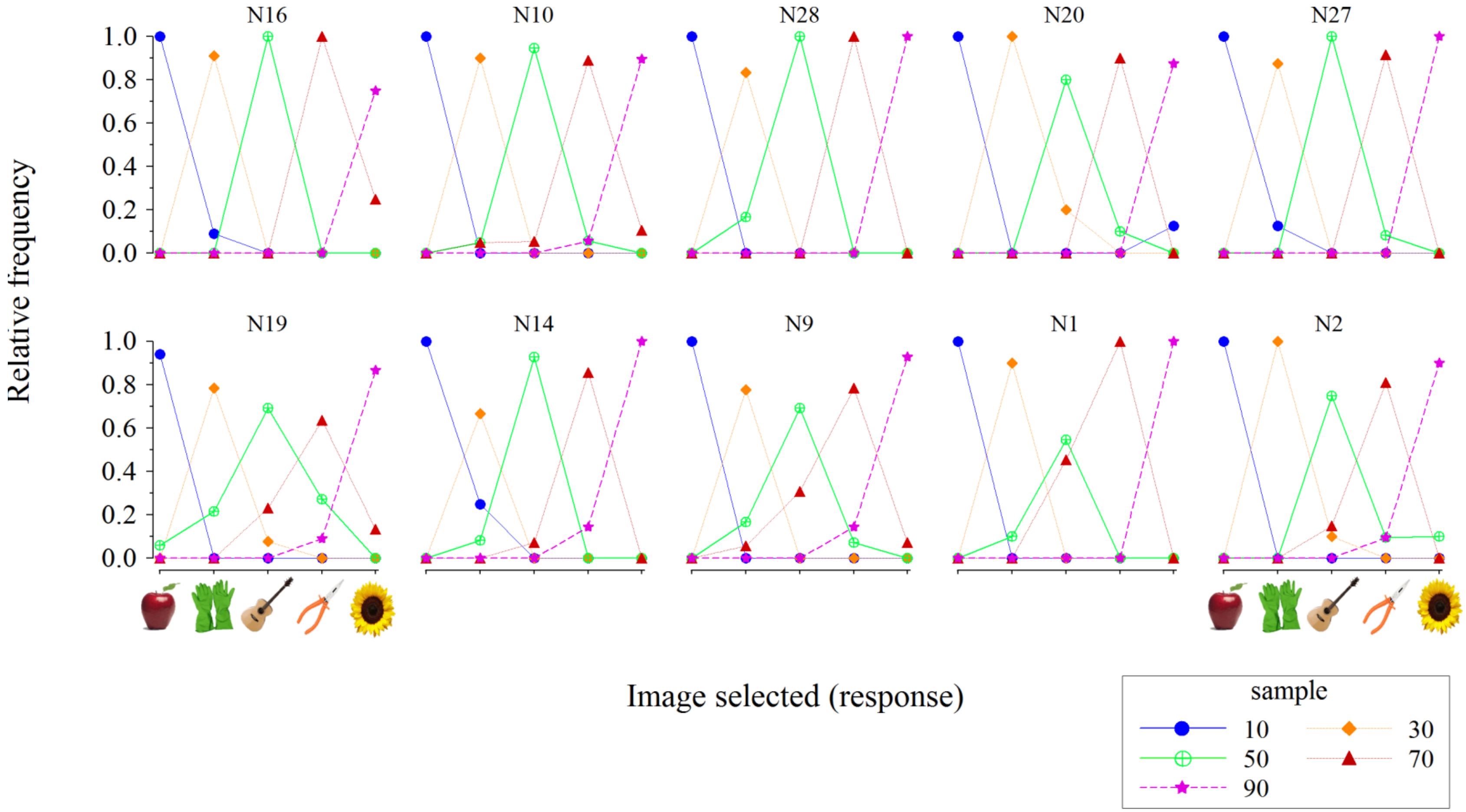


Figure 3. Individual generalization gradients of the 'Perceptual Training' treatment group in the Number dimension.

Time

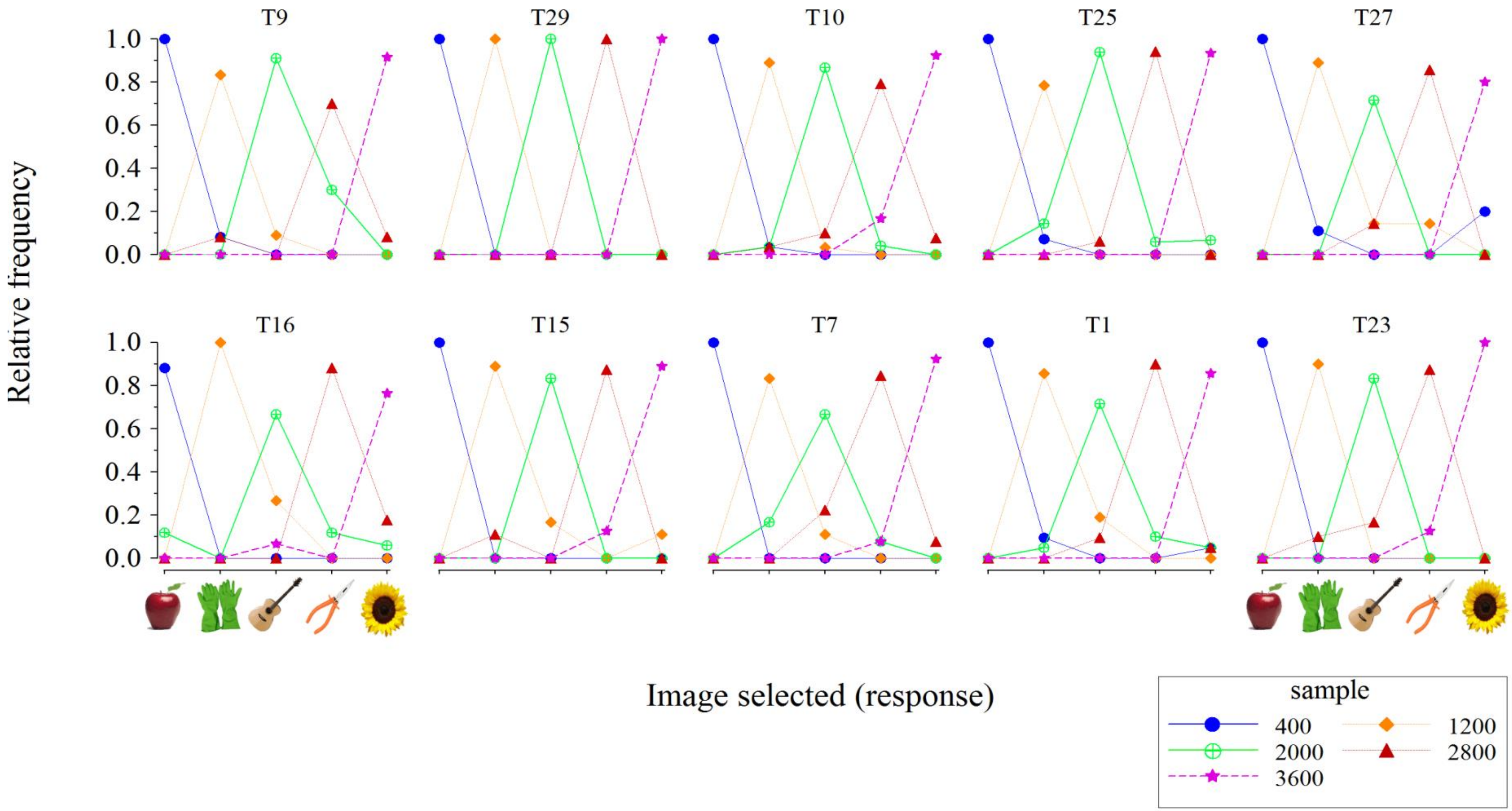


Figure 4. Individual generalization gradients of the 'Perceptual Training' treatment group in the Time dimension.

Performance in the Bisection tasks

General features of temporal and numerical discrimination. All participants produced a psychometric function in which proportion of “Many” (or “Long”) responses increased as a function of the sample’s value (Refer to Figures 7 and 8 for the individual psychometric functions). The graphs in Figure 5 show the groups’ averages of “Many” responses. For each experimental group, responses tended to increase monotonically with number of dots presented, from 0 to 1, revealing good discrimination between the anchor stimuli. This finding, in itself, is of no particular surprise, given that a ratio of 1:9 between the anchor values far outpaces the limits of children and adult humans’ numerical discrimination ability. Such is also the case in regards to humans’ temporal discrimination and, accordingly, the psychometric curves of temporal bisection also increase from 0 to 1, as illustrated in Figure 6.

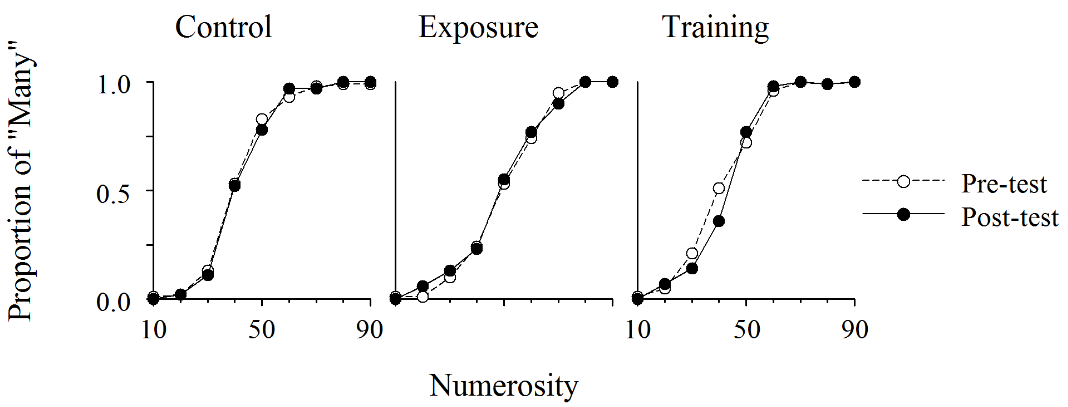


Figure 5. Average proportion of “Many” responses plotted as a function of numerosity (number of dots constituting a set), from the two numerical bisection sessions. The white dots represent the results from the first session (Pre-test), whereas the black dots represent the results from the second session (Post-test).

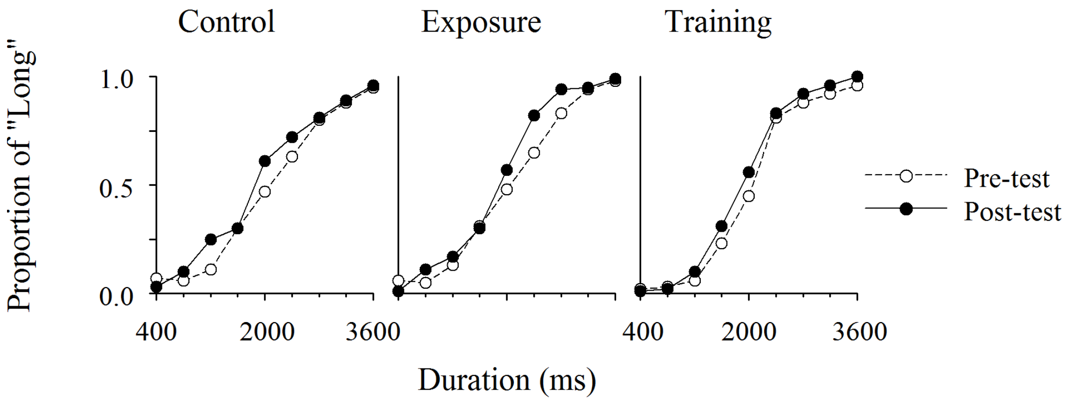


Figure 6. Average proportion of "Long" responses plotted as a function of duration (in milliseconds), from the two temporal bisection sessions. The white dots represent the results from the first session (Pre-test), whereas the black dots represent the results from the second session (Post-test).

Before we address differences between the experimental groups, a brief description of the psychometric curves' properties will follow. The Bisection Point (BP) and Weber Ratio (WR) were estimated by fitting the logistic function from the Origin software (OriginLab, Northampton, MA) to the individual psychometric functions. The logistic equation is: A2 + (A1-A2)/(1 + (x/x0)^p), where A1 and A2 are the initial and final asymptotic values, respectively, x0 is the inflection point (i.e., the Bisection Point), and p is the slope factor/steepness of the curve. All individual psychometric functions were significantly well fitted by the logistic function ($p < .05$), both on the Number dimension's Pre-Test (mean $R^2 = .99$; min. = .96; max. = 1) and Post-Test (mean $R^2 = .99$; min. = .97; max. = 1) sessions, as well as on Time's Pre-Test (mean $R^2 = .97$; min. = .71; max. = 1) and Post-Test (mean $R^2 = .97$; min. = .82; max. = 1) sessions.

Based on the fitted logistic functions we estimated the .25 and .75 quartiles. To estimate the Weber Ratio we divided the difference limen (the difference between the .75 and .25 quartiles) by the Bisection Point. A rule-of-thumb when visually inspecting the psychometric functions is that the steeper the curve the higher the numerical sensitivity, thus the smaller the Weber Ratio. Reversely, flat psychometric functions show low sensitivity to the presented numerosities and are characterized by larger Weber Ratios (for more details about the analysis of psychometric bisection curves, refer to Tavolga, 1969, p.58; Gibbon, 1981; Droit-Volet & Wearden, 2001).

Let us address the Bisection Points and Weber Ratios from the first bisection session (Pre-test), so we can describe adult's numerical and temporal discrimination in a 1:9 ratio, prior to any possible effects of the treatment. First, a question of interest in the literature has been the location of the bisections points. Specifically, studies typically investigate whether the bisection points are located closer to the Arithmetic mean ($AM = (a + b)/2)$) or the Geometric mean ($GM = \sqrt{(a \times b)}$).

In the case of our numerical task, a 10 vs. 90 discrimination, the AM is 50, and the GM is 30. The average of the obtained individual BP was 43.16 (SD = 9.92), a value in-between the AM and the GM. One-sample t-tests found a significant difference between the obtained BP and both the AM ($t(29) = -3.78$, $p = .001$) and GM ($t(29) = 7.27$, $p < .001$). In this case, our data does not fit with either of the explanations that a BP at each mean would entail regarding the format of the subjective scale for numerosities (Cantlon, Cordes, Libertus, & Brannon, 2009, Fig. 1).

As for the temporal 400 vs. 3600 bisection, the AM is 2000 and the GM is 1200. The average of the individual BP, which was 2020.49 (SD = 444.53), was statistically different from the GM ($t(29) = 10.11$, $p < .001$), but not from the AM ($t(29) = .25$, $p = .802$). Thus, we found a BP more similar to the AM, which suggests a linear representation of durations, with constant variability (Allan, 2002; Kopec & Brody, 2010; Droit-Volet & Izaute, 2009).

Effects of the treatments: Number. One way to evaluate the possible effect of the treatment in-between the bisection assessments is to compare the proportion of "Many" responses and of the location of the BP (bias) and the size of the WR (sensitivity), between the first and second bisection sessions (see, e.g., Droit-Volet, Clément, & Fayol, 2008; Droit-Volet & Izaute, 2008). The individual BP and WR values are presented in Table 1. Please refer to the individual scatterplots, depicted in Figure 7, for the empirical functions from which these parameters were estimated.

Proportion of "Many" responses. Visual inspection of the group curves, depicted in Figure 5, first suggests that the Pre- and Post-test curves are quite near each other. The group that seems to present a larger shift between the Pre-and the Post-test is the one that experienced the perceptual training. Visually, the post-test seems to have decreased the proportion of "Many" responses. A mixed model ANOVA was carried with probe numerosity (10 to 90) and test moment (pre- vs. post-test) as the within-group factors, and treatment condition (Control, Exposure, and Training) as the between-group factor. Besides the significant effect of proportion of “Many” responses increasing with probe numerosity ($F(2,27) = 304.136$, $p < .001$), there was no significant effect of test moment ($F(2,27) = .144$, $p = .707$). Moreover, and more important, there was no significant effect of the treatment condition ($F(2,27) = 2.517$, $p = .099$), and neither an interaction effect between test moment and treatment ($F(2,27) = .306$, $p = .739$) or between numerosity and treatment ($F(2,27) = 2.512$, $p = .002$). Thus, the type of treatment in-between bisection moments did not alter the proportion of “Many” responses.

Response bias (Bisection Points). A mixed ANOVA on the Bisection Points obtained in the pre- and post-test assessments found no significant effect of the test moment by itself ($F(2,27) = .537$, $p = .470$). More important, the three experimental groups did not differ significantly ($F(2,27) = 2.458$, $p = .105$), neither there was an interaction effect between treatment groups and the test moment ($F(2,27) = .772$, $p = .472$).

Sensitivity to Number (Weber Ratios). In Table 1, we can observe that the largest mean of individual WR in the Post-Test session belongs to the Control Group ($M = .126$, $SD = .061$). By order of decreasing WR, follows the Exposition group ($M = .088$, $SD = .064$) and, finally, the Training group ($M = .071$, $SD = .055$). The observation that the smaller WR were found in the Training Group, suggests that sensitivity to number increased due to the perceptual training protocol. However, these differences are not statistically significant. In fact, the mixed ANOVA on the WR values found that the three groups did not differ significantly ($F(2,27) = 2.715$, $p = .084$). There was a significant effect of the assessment moment ($F(2,27) = 6.030$, $p = .021$), in that sensitivity to probe numerosity tended to increase in the second assessment (lower WR). Finally, there was no significant interaction effect between test moment and treatment group ($F(2,27) = 2.549$, $p = .097$).

Table 1. Individual Bisection Points (BP) and Weber Ratios (WR) in the Numerical Bisection (10 vs. 90 dots) Pre- and Post-test assessments.

		Age (yrs)	Pre-test		Post-test	
			BP	**WR**	**BP**	**WR**
Control						
	N05	53.19	33.266	0.216	39.507	0.179
	N06	52.85	43.965	0.176	51.446	0.202
	N07	42.4	43.324	0.349	45.801	0.136
	N08	30.79	38.869	0.159	39.798	0.124
	N13	38.65	40.141	0.098	39.17	0.122
	N18	47.73	41.128	0.037	41.128	0.037
	N21	47.6	38.283	0.135	41.938	0.118
	N22	48.97	36.307	0.081	35.969	0.195
	N25	28.99	44.718	0.201	44.443	0.134
	N26	31.44	38.288	0.126	30.856	0.016
	Mean	42.26	39.829	0.158	41.006	0.126
	SD	9.27	3.593	0.087	5.579	0.061
Exposure						
	N03	46.27	41.423	0.095	37.075	0.139
	N04	40.6	65.861	0.082	71.013	0.042
	N11	44.14	53.509	0.094	44.728	0.057
	N12	43.33	47.898	0.158	54.955	0.113
	N15	23.33	55.944	0.084	57.666	0.103
	N17	18.7	56.117	0.055	50.225	0.011
	N23	44.79	25.614	0.134	18.431	0.226
	N24	45.29	40.26	0.015	40.26	0.015
	N29	21.11	42.647	0.085	45.865	0.102
	N30	45.43	64.975	0.062	63.664	0.079
	Mean	37.30	49.425	0.086	48.388	0.088
	SD	11.37	12.379	0.040	14.866	0.064
Training						
	N01	51.86	51.373	0.088	49.836	0.069
	N02	53.31	22.567	0.266	20.125	0.019
	N09	42.56	41.908	0.138	39.509	0.019
	N10	19.94	51.606	0.041	40.733	0.051
	N14	40.68	42.762	0.16	44.716	0.088
	N16	18.52	46.602	0.27	42.996	0.159
	N19	26.08	43.381	0.066	47.785	0.146
	N20	24.39	30.096	0.131	39.798	0.124
	N27	25.28	37.011	0.201	51.3	0.034
	N28	22.19	34.92	0.137	48.744	0.014
	Mean	32.48	40.223	0.150	42.554	0.071
	SD	13.31	9.229	0.077	8.965	0.055

Note. The Arithmetic mean is 50, and the Geometric mean is 30.

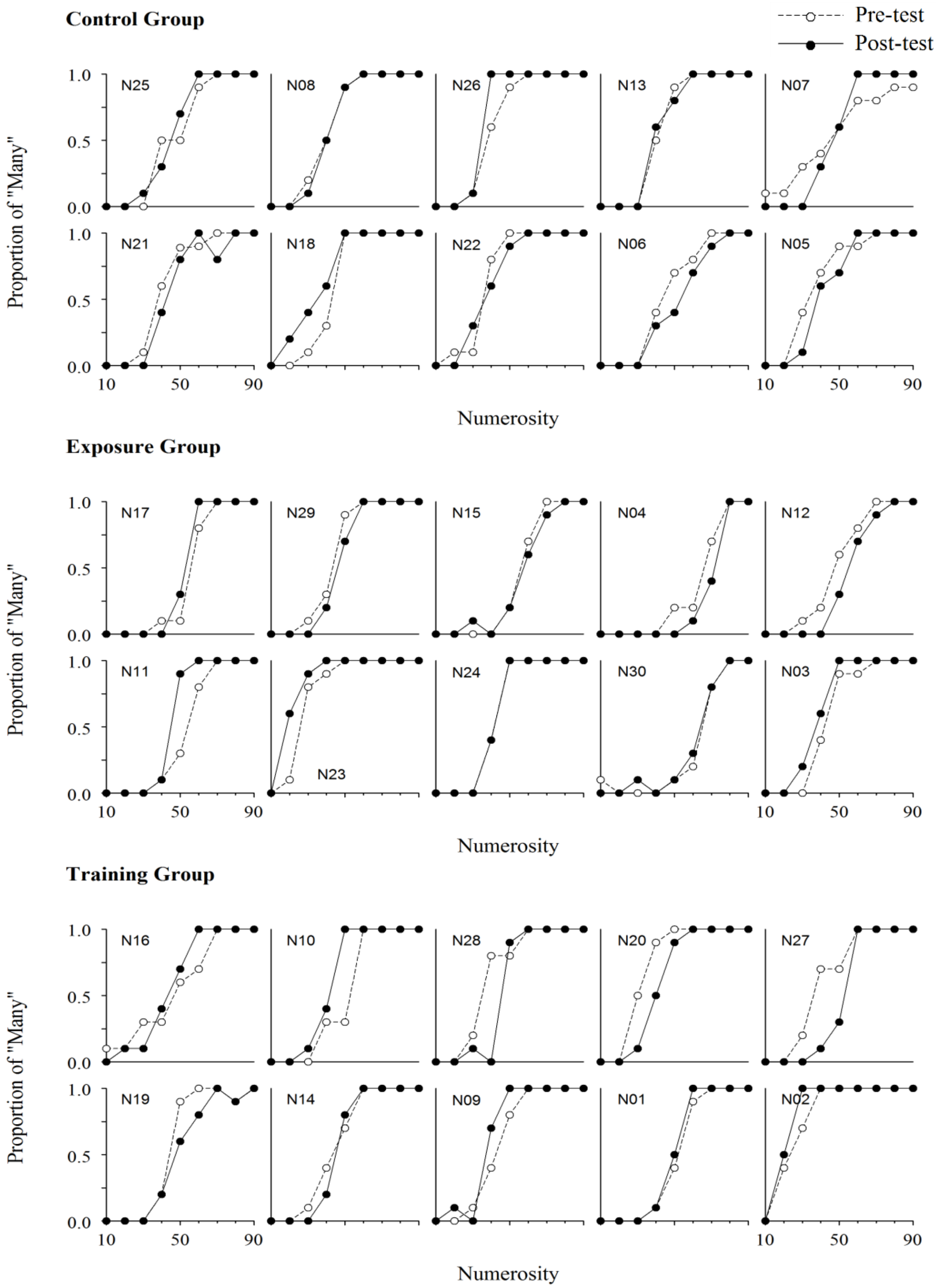


Figure 7. Individual scatterplots of the Numerical Bisection tasks. The white dots represent the mean proportion of "Many" responses during the first bisection task (Pre-test), the filled black dots the proportion of "Many" during the second bisection task (Post-test).

Effects of the treatments: Time. The individual BP and WR values from the Pre- and Post-test Temporal Bisection tasks are presented in Table 2. Please refer to the individual scatterplots, depicted in Figure 8, for the empirical functions from which those parameters were estimated. To test for the effects of probe duration, mixed model ANOVAs were carried with probe duration and test moment (pre- vs. post-test) as the within-group factors, and treatment condition (Control, Exposure, and Training) as the between-group factor.

Proportion of "Long" responses. The only significant effects were that of probe duration ($F(2,27) = 249.066$, $p < .001$) and test moment ($F(2,27) = 9.591$, $p = .005$). As we had observed in Figure 6, proportion of "Long" responses increased as a function of duration in both test moments, but subjects judged the durations as being longer in the Post-test assessment. Yet, the three groups did not differ significantly ($F(2,27) = .067$, $p = .935$; additionally, there was no interaction effect between test moment and treatment group, $F(2,27) = .034$, $p = .967$).

Response bias (BP). Bisection points tended to be statistically significantly smaller in the post-test moment ($F(2,27) = 9.165$, $p = .005$) but, yet again, the three groups did not differ significantly amongst themselves ($F(2,27) = .054$, $p = .948$). There was also no statistically significant interaction effect between test and treatment group ($F(2,27) = .235$, $p = .792$).

Sensitivity (WR). In Table 2, we observe that the larger mean WR values during the Post-test session, belonged to the Control group ($M = .223$, $SD = .153$). By order of decreasing magnitude, came the WR in the Exposition group ($M = .171$, $SD = .077$) and, finally, the smallest WR were found in the Training group ($M = .150$, $SD = .121$). This was similar to the Number dimension, where differences between groups suggested that exposition and, especially, perceptual training increased sensitivity in temporal discriminations. However, as had also been the case with the Number dimension, these differences did not reach statistical significance, Indeed, the mixed-model ANOVA showed that the three groups did not differ significantly on their sensitivity to duration ($F(2,27) = .999$, $p = .381$), neither was an effect of the test moment ($F(2,27) = .039$, $p = .845$), nor an interaction effect between test and treatment ($F(2,27) = .208$, $p = .814$).

Table 2. Individual Bisection Points (BP) and Weber Ratios (WR) in the Temporal Bisection (400 vs. 3600 ms) Pre- and Post-test assessments.

		Age (yrs)	Pre-test		Post-test	
			BP	**WR**	**BP**	**WR**
Control						
	T03	38.64	2721	0.071	3236	0.045
	T04	38.1	1371	0.488	1680	0.326
	T05	48.85	2634	0.116	2478	0.183
	T06	40.49	1742	0.110	1324	0.206
	T13	18.66	1835	0.398	1323	0.573
	T14	18.61	1927	0.178	1775	0.093
	T19	41.43	1872	0.257	1938	0.098
	T20	18.83	2122	0.226	1727	0.159
	T21	53.04	1446	0.215	1045	0.237
	T22	49.41	2886	0.250	2436	0.313
	Mean	39.61	2055.53	0.231	1896.21	0.223
	SD	13.29	527.91	0.130	658.05	0.153
Exposure						
	T02	20.56	1418	0.355	1693	0.250
	T08	37.96	1185	0.027	733	0.250
	T17	40.74	2691	0.199	2077	0.124
	T18	33.88	2532	0.096	2264	0.150
	T11	38.54	1596	0.333	1205	0.303
	T12	18.85	2323	0.127	2056	0.074
	T26	33.38	1642	0.181	1676	0.138
	T28	33.91	1892	0.078	1855	0.136
	T24	19.17	2468	0.159	2254	0.204
	T30	29.31	2461	0.033	2238	0.084
	Mean	30.63	2020.67	0.159	1805.02	0.171
	SD	8.30	537.29	0.113	502.94	0.077
Training						
	T01	42.38	1688	0.621	1802	0.445
	T07	40.76	2019	0.029	1744	0.069
	T15	39.98	2216	0.128	1998	0.156
	T16	37.12	2079	0.113	2144	0.127
	T09	18.34	2339	0.168	2065	0.148
	T10	24.54	2146	0.092	2191	0.076
	T25	30.09	2107	0.070	1858	0.133
	T27	36.54	1477	0.114	1512	0.080
	T23	43.7	1945	0.209	1201	0.022
	T29	18.69	1836	0.121	2095	0.248
	Mean	33.21	1985.26	0.166	1861.04	0.150
	SD	9.66	258.11	0.167	312.29	0.121

Note. The Arithmetic mean is 2000 milliseconds, and the Geometric mean is 1200 ms.

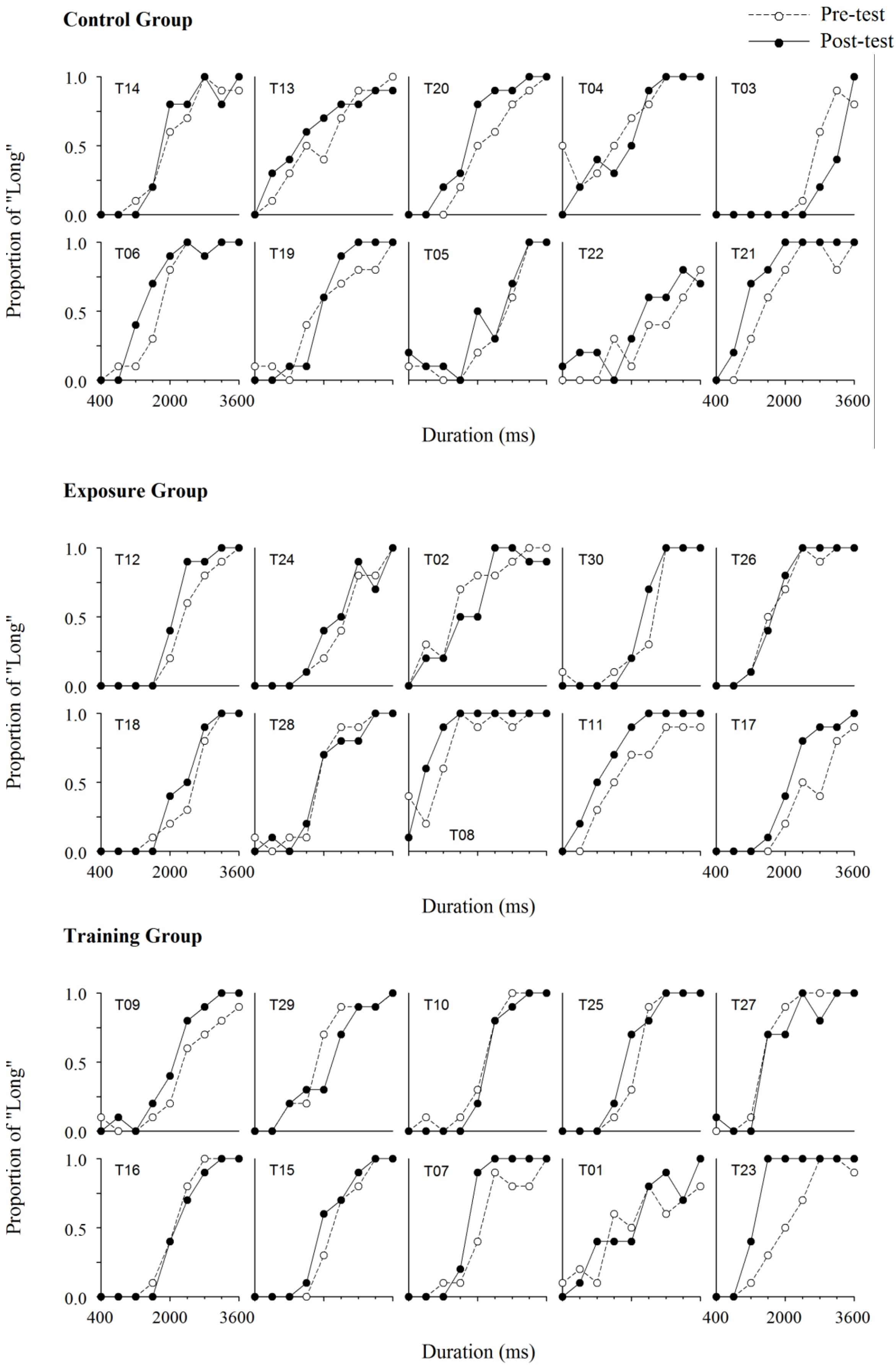


Figure 8. Individual scatterplots of the Temporal Bisection tasks.

Conclusions

The current study aimed to investigate how sensitivity to temporal and numerical stimuli is improved by practice and, additionally, whether a mere exposition to the test stimuli or a discriminative training produced differentiated degrees of amelioration. Namely, we contrasted human performance in temporal and numerical bisection tasks, which either anteceded or followed a discriminative training or simply the presentation of the relevant stimuli. Both group and individual performance in the Post-test phase (2^{nd} bisection session) was contrasted to the Pre-test (1^{st} bisection), by inspecting the psychometric curves of Proportion of “Long” (or “Many”) responses as a function of probe duration (or numerosity). We were particularly interested in seeing whether the three treatment conditions differentially affected sensitivity (Weber Ratio) and response bias (Bisection Point).

The between-groups comparison of the mean Weber Ratios in the Post-test phase showed that sensitivity was lower in the Control group. Sensitivity increased in the Exposure group and was highest in the Training group. This suggested that mere exposure and, more so, categorical training affected numerical and temporal sensitivity. However, the inferential statistics tests did not confirm the statistical significance of these differences.

It is possible that the experimental treatments failed to affect performance due to the procedural and sample features, such as the age of our participants and, above all, the large (1:9) ratio between the two anchor stimuli. To our knowledge only one numerical bisection study with human adults has tested them in a ratio larger than ours (1:20 in Tan & Grace, 2012). Previous number bisection studies with nonverbal numerosities have tested adults with a 1:4 ratio (Roitman, Brannon, Andrews, & Platt, 2007), or a 2:5 ratio (Droit-Volet, Clement, & Fayol, 2008; Droit-Volet, 2010). Our ratio, 1:9, is well-above the limits of adults’ acuity in nonverbal numerical discrimination (Halberda & Feigenson, 2008), as well as temporal discrimination (e.g., Zéltanti & Droit-Volet, 2012; Droit-Volet & Zélanti, 2013; Droit-Volet, Tourret, & Wearden, 2004). We wonder whether testing adults in an easy-to-discriminate ratio is related to the fact that many participants’ individual psychometric functions rise in a step-like fashion, instead of gradually as a function of numerosity (or duration). Perhaps our adults’ bisection performance was also the result of “all-or-none” rules of responding such as «if sample is not exactly the ‘few’ (or ‘short’) value, then answer “many” (or ‘long’)». Had the discrimination been harder (e.g., smaller ratio such as 2:3), would these step-like curves ensued as well? And if not, would responding be more dependent on perceptual features and less so in mediating rules of responding? Unfortunately, inspection of individual psychometric functions is mostly absent in human numerical and temporal bisection studies and, as such, we cannot verify whether step-like (“all or nothing”) type of responding is affected by the small/large anchor ratio.

In conclusion, future studies ought to investigate the effect of the perceptual treatment protocols (e.g., Angulo & Alonso, 2012) in bisection discriminations between smaller ratio values. Even with the larger 1:9 ratio we applied, differences between groups hinted at an effect of exposition and categorical training. But sample size as well as sampling method could have decreased our chances of uncovering a true and significant difference between the groups. In order to boost the statistical power of a future study, we would thus increase the number of trials per probe numerosity/duration or, in alternative, recruit more participants to each experimental group.