DOCTORAL (PHD) DISSERTATION

Petia Kojouharova

The Numerical Distance and Size Effects in Symbolic Numerical Cognition

2019

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF EDUCATION AND PSYCHOLOGY

Petia Kojouharova

The Numerical Distance and Size Effects in Symbolic Numerical Cognition

DOCTORAL SCHOOL OF PSYCHOLOGY
Head of the Doctoral School: Zsolt Demetrovics, PhD, Dsc

COGNITIVE PSYCHOLOGY PROGRAMME
Head of the Program: Ildikó Király, PhD
Supervisor: Attila Krajcsi, PhD

Committee

Chair:	Prof. Márk Molnár, PhD, DSc
Secretary:	Anett Ragó, PhD
Opponent:	Márta Volosin, PhD
Opponent:	Bert Reynvoet, PhD
Members:	Kristóf Kovács, PhD
	Christophe Heintz, PhD
	Ernő Téglás, PhD
	Dorottya Ujfalussy, PhD

Budapest, 2019

Acknowledgments

First and foremost, I would like to thank my supervisor, Attila Krajcsi, for his guidance, for all the knowledge he shared, and for the professional and emotional support. I am also grateful to the members of The NumberWorks Lab, to Gábor Lengyel and Lilla Hodossy for all the collaboration from the very start, and also, more recently, to Ákos Laczkó, Orsolya Kis, and Tamás Szűcs. It has been a privilege to have worked together on these novel and exciting projects, a part of which are the studies in this thesis.

I would like to thank the members of the Cognitive Psychology Research Group at the Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences: Zsófia Anna Gaál and István Czigler for their support and valuable advice during the writing of the thesis as well as for the opportunity to work among peers in a great research environment. I am also grateful to István Sulykos, Boglárka Nagy, Domonkos File, Zsuzsanna D'Albini, Flóra Bodnár, Petra Csizmadia, and Béla Petró for all their help.

I would like to thank Nóra Gaál for the safe (and flexible) work environment, for her support, and for her kindness. My thanks also go to the International Office at the Faculty of Education and Psychology, Eötvös Loránd University for having been great coworkers, and to the Dean, Zsolt Demetrovics, without whose support I would not have been able to start the PhD programme.

I would like to thank the Cognitive Psychology Department at the Faculty of Education and Psychology, Eötvös Loránd University for their efforts to create and maintain a research community, and especially Ildikó Király for her professional advice as well as the opportunity to "look around", and Anett Ragó for the professional and emotional help.

Finally, I would like to thank my life partner, Tibor Drávai, for his unwavering love, faith, and support, my parents for always being there for me, and my friends Renáta Kiss, Ralitsa Hiteva, Denitsa Genevska, and Denitsa Yalamova for all the conversations and all the booze.

Thank you!

Table of Contents

List of the Publications Included in the Thesis

Krajcsi, A., Lengyel, G., & **Kojouharova, P.** (2016). The source of the symbolic numerical distance and size effects. *Frontiers in psychology, 7*, 1795. https://doi.org/10.3389/fpsyg.2016.01795

Krajcsi, A., & **Kojouharova, P.** (2017). Symbolic numerical distance effect does not reflect the difference between numbers. *Frontiers in psychology, 8*, 2013. https://doi.org/10.3389/fpsyg.2017.02013

Kojouharova, P., & Krajcsi, A. (2018). The Indo-Arabic distance effect originates in the response statistics of the task. *Psychological research*, 1-13. https://doi.org/10.1007/s00426-018-1052-1

Kojouharova, P., & Krajcsi, A. (2019). Two components of the Indo-Arabic numerical size effect. *Acta psychologica, 192*, 163-171. https://doi.org/10.1016/j.actpsy.2018.11.009

Abstract

When two numbers have to be compared based on their numerical value (i.e., participants must respond to the “which one is larger?” question), two phenomena are observed. The participants respond faster and with fewer errors when the difference between the numbers is large which is known as the distance effect. Their performance is also better when the numbers are smaller which is termed the size effect. Both effects are considered to be a consequence of a ratio-based representation of numbers: An innate, continuous, analogue representation of quantities that is shared across species, termed the Analogue Number System (ANS) which works according to Weber's law. This same system is thought to be the mechanism behind both symbolic and non-symbolic numerical processing. However, there is a plausible alternative account for the sources of the two effects, and thus for symbolic numerical processing. According to the Discrete Semantic System (DSS) account the numbers could be represented as discrete nodes, the distance effect could stem from the strength of the connections between the nodes, and the size effect could be rooted in the frequency of the numbers. In four studies we systematically investigated the sources of the distance and size effects in both new, artificial numbers and Indo-Arabic numbers. In the first and the fourth study the frequency of the numbers was manipulated to induce changes in the size effect. In the second and third study the associations of the numbers with the “small-large” properties were manipulated to see whether the distance effect would be modified. The results showed that: 1) the source of the distance effect are the associations of the numbers with the “small-large” properties, 2) the distance effect is flexible, 3) the size effect is rooted in the frequency of the numbers and is only partially flexible, 4) the two effects dissociate. Overall, the results for symbolic numbers were inconsistent the ANS account, but in line with the DSS account. Thus, the DSS provides a better explanation for symbolic number comparison. The present results are also in line with recent findings in the field that support the existence of a separate system for processing symbolic numbers.

Absztrakt

Amikor az értékük alapján kell összehasonlítani két számot (vagyis arra a kérdésre választ adni, hogy „melyik a nagyobb?”), két jelenség figyelhető meg. A résztvevők gyorsabban adnak választ és kevesebbet hibáznak, amikor a számok közötti különbség nagyobb, amely jelenség úgy ismert, hogy távolsághatás. Emellett, a teljesítményük jobb, amikor a számok kisebbek, amit nagysághatásnak neveznek. Mindkét hatást a számok arány alapú reprezentációjának a következményének gondolják: Egy veleszületett, folytonos, analóg mennyiség reprezentáció, az úgy nevezett Analóg Mennyiség Rendszer (AMR), amely állatok esetében is megtalálható, és a Weber-törvény szerint működik. Ezt a rendszert vélik a szimbolikus és nem szimbolikus numerikus feldolgozás mögötti mechanizmusnak. Azonban létezik egy lehetséges alternatív modell a két hatás forrására, és így a szimbolikus numerikus feldolgozást illetően. A Diszkrét Szemantikus Rendszer (DSZR) szerint a számok diszkrét csomópontokként vannak reprezentálva, a távolsághatás a számok közötti kapcsolatok erősségéből származhat, a nagysághatás pedig a számok gyakoriságából. Négy kutatásban szisztematikusan vizsgáltuk meg a távolsághatás és a nagysághatás forrásait új, mesterséges számokkal és indo-arab számokkal. Az első és a negyedik kutatásban a számok gyakoriságát manipuláltuk ahhoz, hogy megváltoztassuk a nagysághatást. A második és a harmadik kutatásban a számok asszociációit a „kicsi-nagy” tulajdonságokkal manipuláltuk ahhoz, hogy azt vizsgáljuk meg, hogy változik-e a távolsághatás. Az eredmények azt mutatták, hogy: 1) a távolsághatás forrása a számok és a „kicsi-nagy” tulajdonságok közötti asszociációk, 2) a távolsághatás rugalmas, 3) a nagysághatás a számok gyakoriságából származik és csak részben rugalmas, 4) a két hatás disszociál. Összességében, a szimbolikus számok esetében kapott eredmények inkonzisztensek az AMR modellel, de összhangban vannak a DSZR modellel. Így a DSZR jobb magyarázatot ad a szimbolikus numerikus összehasonlításra. A jelenlegi eredmények összhangban vannak a szakirodalom újabb eredményeivel, amelyek támogatják a külön rendszer létezését a szimbolikus numerikus feldolgozás esetében.

Introduction

If a human ability, especially one that has been known to be highly dependent on language and culture, shows signs of being present early in evolution (other species show at least a rudimentary form of it) and/or early in development (infants are able to solve a task that tests this ability), then a major question becomes recognizing what part of the ability is due to innate mechanisms and what part depends on other factors, including the role of culture in general and language specifically. Further issues that arise are whether the more sophisticated, language-based knowledge builds on the innate knowledge, whether there is an interaction between them, whether any other mechanisms are involved, or whether the mechanisms are mostly independent of each other.

Handling quantities – estimation, comparison, counting – is one such ability, and forms a part of the research field of numerical cognition. Furthermore, it is an ability that is indispensable in the everyday life, and can have a great impact on decisions, on the quality of life etc. A better understanding of its mechanisms could help with teaching mathematics more efficiently and intervening in related disorders such as dyscalculia.

Non-symbolic (e.g., handling sets of objects, sound sequences etc.) and symbolic (e.g., handling Indo-Arabic digits, number words etc.) numerical processing is studied via phenomena that can be observed in tasks requiring basic arithmetic skills such as quantity estimation or comparison, and probably the most widely used paradigm among those is the number comparison task in which the participants have to choose the larger (smaller) of two numbers. In this task two of the effects observed with both nonsymbolic and symbolic numbers are the distance effect and the size effect. The distance effect means that when comparing two numbers based on their numerical value, responses are faster and errors are fewer when the numerical distance between the numbers is larger. The size effect means that performance is better when comparing smaller numbers. Both effects are interpreted in the currently prevalent model of numerical cognition as the indicators of an innate, continuous, noisy representation system termed the Analogue Number System (also Approximate Number System, ANS) that works according to Weber's law, i.e., depends on the ratio of the two quantities. Thus, the conclusion drawn about our numerical abilities is that the source for both effects is the ratio of the two numbers to be compared.

This is likely true for non-symbolic comparison. All research up until now shows that animals can compare and estimate quantities in a way predicted by the ANS, and infants show similar abilities very early in life. However, in the case of symbolic numbers there is a plausible alternative. In a semantic network the distance effect could reflect the strength of the connections between the numbers (which are nodes in that network) or between the numbers and certain properties such as the "small-large" properties. The size effect could simply be a frequency effect as in that smaller numbers are more frequent and thus easier to process. In other words, we propose that the two effects have different sources and are independent of each other in symbolic numerical processing. The alternative model proposed as the mechanism behind symbolic numerical cognition is the Discrete Semantic System (DSS), which supposes a semantic network or a mental-lexicon-like representation for symbolic numbers.

In the thesis I first describe the two accounts for symbolic numerical cognition, the importance of the distance and the size effects, and the aims of the studies. Then, I briefly describe the studies and the methods that we applied to differentiate between the two accounts. The four Thesis Studies follow, in which we first showed that it is impossible to differentiate between ANS and DSS accounts in the case of symbolic numbers if they are compared directly. Then, we systematically manipulated the associations between the numbers and the "small-large" properties as well as the numbers' frequency for both artificial numbers and Indo-Arabic numbers, and we showed that this manipulation leads to behavioral changes (i. e., a change in the distance and size effects) that is inconsistent with the ANS explanation, but very much consistent with the DSS account. Finally, I shall discuss that we demonstrated that the symbolic numerical abilities are independent from, or at the very least have a qualitatively different relationship with, the innate mechanism for handling quantities. This has both practical implications as well as implications about the relationship between innate abilities and language-based abilities.

Theoretical Background

An innate representation.

In a slight departure from the usual route for introducing numerical cognition in more recent years, I would like to start with a rather fascinating paper on the representation of numbers by Galton (1880). He observed that some adults and many children represent numbers in a visual form, in space and often in the form of a line. The paper already describes some of the properties investigated since then, e.g., an ordered, spatially-oriented from left to right number line, the overrepresentation of special numbers such as 10s, 100s etc., a division between smaller and larger numbers, language-dependent specificity such as the special places of the numbers 12 and 20 in the English language, a possible dissimilarity between the verbal and the numerical forms. In Galton's interpretation this topical representation is a mnemonic technique for memorizing numbers. More importantly, he links it to individual inclinations in childhood as if children search for the appropriate way to match an internal representation to language.

Galton's work could be seen as a starting point that could lead us to the formulation of the following questions: Is there an innate representation of numbers, what is its connection to language, and how can both be explored and measured? The idea of an innate representation introduces the option that numerosity is a property of real world objects similar to luminance, loudness, weight, and is perceived and represented in a similar manner. If it is, then numbers, which are values of that property, should cause similar behavior in tasks of relative judgment, and their representation should be described by psychophysical laws such as Weber's law (Fechner, 1860/1912). Weber's law states that the difference between two values of a property has to be sufficiently large to be noticed and the larger the values are, the larger the difference should be. The reason behind it is the overlap between the representations of the values to compare. A highly individual constant calculated as the ratio of the two values and called the Weber fraction defines how well tuned their representations are. More specific quantification of innate representations has been attempted for empirical data, and collections such Crossman (1955) who treats Weber's law as a private case of a more general rule and Welford (1960) provide numerous examples.

The distance effect and the size effect.

The supposition that the representation of numbers could be subject to psychophysical laws was the starting point of Moyer and Landauer's (1967) seminal paper in *Nature*: The study investigated the question of how people judge numbers depending on their values. Two possibilities were put forward: (1) comparison is similar to comparing physical properties and works the same way perception does, or (2) it happens on a higher, "more cognitive" level. The time required for comparing two numbers ("which one is larger") was measured. If the former possibility was true, the closer the numbers were, the more difficult the comparison would be because of the overlapping representations. If the latter possibility was true, the farther away the numbers were from each other, the more difficult the comparison should be as it would take time to reach the representation of the next number. The reaction time and the distance between the numbers correlated negatively. Several psychophysics formulas were tested on the data (based on Welford, 1960), and one of them – $RT = K \times \log(large_{number} / (large_{number} - small_{number}))$ – was deemed a reasonable fit. In other words, the reaction time data was better explained by ratio and not absolute difference. The authors thus concluded that the representation of numbers has psychophysical properties and as such will behave like any other perceptual system, e.g., the visual system.

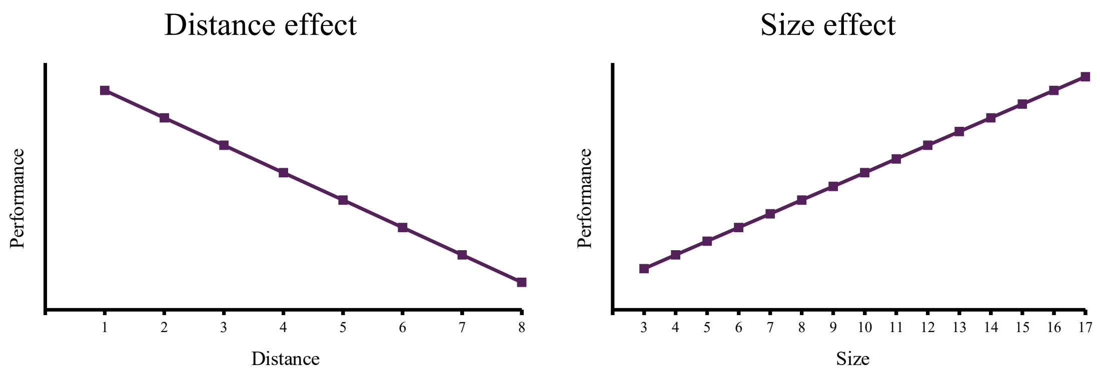


Figure 1. The distance effect (left panel) shows worse performance with smaller distance between the numbers. The x-axis shows the absolute difference. The size effect (right panel) is worse performance with larger numbers. The x-axis shows the effect expressed as the sum of the two numbers. Performance, shown on the y-axis, is measured with error rate or reaction time

Two essential points from the Moyer and Landauer paper should be emphasized, especially in relation with the present thesis. First of all, it introduces the two most important phenomena investigated in numerical cognition that are also the subject here: The comparison distance effect and the comparison size effect in symbolic numbers[1]. The distance effect means that performance in number comparison tasks increases with the increase in the numerical distance between the numbers, whereas the size effect means that performance is worse for larger numbers (Figure 1). Both effects have been replicated in children (e.g., Sekuler & Mierkiewicz, 1977), with number words (e.g., Dehaene & Akhavein, 1995), and in other notations, e.g., Kanji, artificial numbers (Cohen Kadosh, Soskic, Iuculano, Kanai, & Walsh, 2010; Razpurker-Apfeld & Koriat, 2006; Takahashi & Green, 1983). Second, it suggested a frame for interpreting the two effects and a route for further research by linking the representation of numerosity to perceptual-like representations.

The Analogue Number System.

Developmental and comparative research provide evidence for the existence of preverbal, innate mechanisms, and the case is no different for numerical cognition (e.g., Beran, 2008; Beran, Johnson-Pynn, & Ready, 2011; Dehaene, Izard, Spelke, & Pica, 2008; Wynn, 1998). Two such core systems seem to emerge (Feigenson, Dehaene, & Spelke, 2004): A noisy, approximate representation of large numbers that represents numbers either in a logarithmic fashion with fixed variability or linearly with scalar variability, and an exact system for small numerosities (1 to 3 or 4). The latter shows characteristics that are very different from what was found for symbolic numbers, and as such is not subject to further discussion here. The former, however, exhibits the properties of a perceptual-like system, and became a plausible candidate for explaining both non-symbolic and symbolic numerical cognition. This account is termed the Analogue Number System (ANS), and it states that both non-symbolic (e.g., sets of dots) and symbolic (e.g., Indo-Arabic numbers) numbers rely on the same innate, continuous, noisy mechanism for understanding and handling quantity that is shared

1 The distance and the size effects are not the only effects in numerical cognition that are investigated to understand the representation of numerosity. Other phenomena include, for example, the SNARC effect (Spatial-Numerical Association of Response Codes) which shows that small numbers are responded faster to with the left hand and large numbers with the right hand (Dehaene, Bossini, & Giraux, 1993), or the numerical size-congruity effect which shows that the value of a number interferes with its physical size (Henik & Tzelgov, 1982). These effects are not described in detail here, but some will be given as further evidence for our claim.

across species (Cantlon, Platt, & Brannon, 2009; Moyer & Landauer, 1967; Walsh, 2003). The ANS works according the Weber's law, and thus performance in comparison tasks depends on the ratio of the numbers to compare. A brief mention is warranted for a possible extension, the ATOM (A Theory Of Magnitude) model (Bueti & Walsh, 2009; Walsh, 2003), a common magnitude system which represents not only quantity, but also other concepts such as space and time that share features and thus partially overlap on neurological level. Certain features are only linked to one of the magnitude representations, and that is usually observable in interactions between different representations, e.g., physical size and numerosity. More recent developments even point in the direction of numerosity based on the processing of visual features such as total area, circumference, convex hull, frequency differences instead of a specialized mechanism devoted only to numerosity (Dakin, Tibber, Greenwood, Kingdom, & Morgan, 2011; Gebuis & Reynvoet, 2012a, 2012b) or to it being an emergent property of the visual system (Stoianov & Zorzi, 2012) (for a fascinating and rather heated debate on the topic as well as for a novel proposition see Leibovich, Katzin, Harel, & Henik, 2017). Importantly, as all of the above bear the same properties as a perceptual system, either is a good account for a mechanism in which numerical knowledge is rooted in Moyer and Landauer’s vein of thought.

Numerous animal studies support the existence of the ANS. Trained rats were able to press a lever approximately for the required number of times with their error (standard deviation of the response distribution) being proportional to the value of the number (Platt & Johnson, 1971). In another study rats transferred their training on number of noise bursts to sets of noise bursts and cutaneous shocks, i. e., it was not the modality but the number of events that mattered (Meck & Church, 1983). Beran (2012) showed that chimpanzees were able to compare food sets based on the noise made by added items, and also compare a set available visually (they saw the tray with the food items) to a set presented auditorily. In all cases, performance was subject to the ratio of the two numerosities. Even more convincingly, Nieder (2005) found neurons in the prefrontal cortex and in the intraparietal sulcus of primates that fired selectively to sets of dots, with the evidence suggesting that numerical information was first processed in the parietal region. The tuning curves of the neurons had a pattern that mirrored a noisy representation. (For an extensive review on studies with animal subjects see Dehaene, 2011; Feigenson et al., 2004).

Developmental studies furthered the case for a preverbal system. Xu and Spelke (2000) showed that infants can discriminate between 8 and 16 dots in a habituation study, and they could do the same for auditory stimuli (Lipton & Spelke, 2003). The ratio necessary for successful discrimination of numerosities decreased with age. Infants are also able to match numbers crossmodally as observed in a study by Izard et al. (Izard, Sann, Spelke, & Streri, 2009) in which infants were trained on auditory sequences of 4 or 12 sounds and 6 or 18 sounds, and then the sequences and visual stimuli (sets of objects) were presented simultaneously – the infants looked longer at the picture that matched the sequence.

A further source of evidence supporting the existence of a preverbal system is the numerical knowledge of tribes isolated from Western-type civilizations. The Munduruku tribe (Pica, Lemer, Izard, & Dehaene, 2004) performed similarly to French controls in non-symbolic numerical tasks, including showing a comparable Weber fraction: 0.12 for French controls, 0.17 for the Munduruku participants. The Piraha tribe showed similar performance: Weber fraction of 0.15 (Gordon, 2004). Neither of the tribes has words for quantities beyond four or five, and even those number words do not always represent exact quantities; nevertheless, there is a system that works for approximate estimation and comparison.

To sum up, there is a preverbal system that can account for the distance and size effects in both non-symbolic and symbolic numbers (Dehaene, 1992). Furthermore, the distance and the size effects in comparison tasks are considered indicators of the ANS as both can be a consequence of Weber's law. In the ANS model the distance effect is explained as the extent of the overlap between the noisy representations of the numbers – the closer the numbers are, the larger the overlap is, making it more difficult to differentiate the two numbers. The same mechanism also results in the size effect – the larger the numbers are, the larger the overlap is. This account supposes that both effects stem from the same source – the ratio of the numbers.

Studies that questioned the ANS for symbolic numbers.

According to the ANS account for symbolic numbers, they activate this noisy representation automatically. The notations themselves are modality-dependent and can have properties that are separate from the ANS (e.g., semantic properties such as parity, see the triple-code model in Dehaene & Akhavein, 1995), but numerosity itself as well as all related phenomena are grounded in the ANS.

The ANS seems to be a sensible model to explain numerical cognition when symbolic tools are not (yet) available and also as a mechanism that guides symbolic numerical processing. Recent findings, however, question whether the ANS is necessarily activated in tasks with symbolic numbers, and whether the observed effects reflect the same effects found in tasks with non-symbolic numbers. Cohen (2009) showed that the distance effect obtained in a same/different task (a number had to be compared to a standard, e.g., “is 6 the same as 5”) with Indo-Arabic numbers could be described better by the physical similarities between the digits than by their numerical value. Studies that explore notations other than the Indo-Arabic have demonstrated that some effects appear differently, for example, in number words (Dehaene & Akhavein, 1995), or depending on the utilized task (Ganor-Stern & Tzelgov, 2008; Ito & Hatta, 2003). On a different note, Krajcsi (2016) showed that whereas the distance and size effects correlate highly in non-symbolic numbers, they do not correlate in symbolic numbers which suggests a common source in the former but not in the latter case.

Another popular research area for investigating the relationship between ANS and symbolic numerical cognition has been correlating performance in non-symbolic and symbolic numerical tasks with math achievement in children as well as the study of developmental dyscalculia. For example, Halberda et al. (2008) showed that ANS acuity (i.e., the value of the individual Weber fraction) and math achievement correlate in 14-year-old children. Other studies, however, have shown that non-symbolic performance and math achievement do not correlate, whereas symbolic performance and math achievement do (Holloway & Ansari, 2009; Sasanguie, Defever, Maertens, & Reynvoet, 2014), or that early symbolic performance predicts later non-symbolic performance, but early non-symbolic performance does not predict later symbolic performance (Mussolin, Nys, Content, & Leybaert, 2014). Similarly, children with developmental dyscalculia started performing worse in non-symbolic tasks later in life, at 9-10 years of age, whereas for younger children performance was worse only in symbolic tasks (Noël & Rousselle, 2011). An extensive meta-analysis by Schneider and colleagues (2017) suggests that non-symbolic comparison correlates much less with math achievement, but a correlation with the number comparison task was found repeatedly.

The neurological evidence also seems problematic. Piazza and colleagues (2007) found distance-dependent activation (fMRI study, BOLD signal) in the horizontal intraparietal sulcus for both non-symbolic and symbolic numerosities, but later studies (Bulthé, De Smedt, & Op de Beeck, 2014; Bulthé, De Smedt, & Op de Beeck, 2015;

Damarla & Just, 2013) that applied more sensitive methods found only notation-dependent activation. For example, in Bulthé et al.'s (2015) study the activation for one Indo-Arabic digit was more similar to that of one dot (one visual object) than the activation to multiple dots.

Regarding the specificity of the effect, a distance effect similar the one in the number comparison task has been found in ordered non-numerical sequences such as letters (e.g., Jou, 2003; Razpurker-Apfeld & Koriat, 2006; Van Opstal, Gevers, De Moor, & Verguts, 2008), months (Seymour, 1980), artificial symbols (Tzelgov, Yehene, Kotler, & Alon, 2000). Moreover, a distance effect has been found for categorical non-ordered stimuli: In a picture-naming task participants were slower when they had to name an object that followed another object from a close category than when the preceding object was from a far category (Vigliocco, Vinson, Damian, & Levelt, 2002).

Alternative models.

One model that has been proposed as an alternative is the delta-rule connectionist model of Verguts and colleagues (Verguts & Fias, 2004; Verguts, Fias, & Stevens, 2005; Verguts & Van Opstal, 2014). In this model the values of the numbers are represented as nodes in a hidden layer between the input (the number in a specified notation) and the output (a response as required by the task). The activation of a node also produces noise with fixed width which can be interpreted as activation spreading to the neighboring nodes. The distance effect obtained in a comparison task in this model is explained by the connection weights between the value nodes and the response nodes of "smaller" and "larger". The size effect in this model can be achieved by introducing biased frequency such as the numbers' everyday frequency (Dehaene & Mehler, 1992). The delta-rule connectionist model presents a good account not only for the distance and size effects, but can also explain the same effects in non-numerical ordered sequences and can account for other numerical effects such as the priming distance effect.

In a similar vein, albeit via a different approach, Henik and Tzelgov (1982) suggested that some basic elements (primitives) are stored in the long term memory, e.g., integers from 1 to 9 and the number 0 (Pinhas & Tzelgov, 2012), while others are not. The basic elements or primitives can be imagined as the nodes in a semantic system that represent the values of the numbers and which can be combined to form other numbers.

The Discrete Semantic System.

Our alternative proposal is the Discrete Semantic System (DSS) which was first proposed in Krajcsi (2016) and Krajcsi, Lengyel, and Kojouharova (2016), the first study of this thesis. The DSS works similarly to the mental lexicon or a semantic network and is capable of processing symbols and abstract concepts. In this model numbers are stored as nodes in a network, and the effects observable in different tasks depend on the strength of their semantic relations to other nodes, i.e., on the connection weights. The DSS can produce the comparison distance effect, and there are at least two possible sources. Close meaning, i.e., close values, may also mean stronger connections between the nodes, and spreading activation from one of the nodes (numbers) activates the rest to some extent depending on the strength of the connection. The distance effect could also be rooted in the numbers' associations with the "small" or "large" properties where, in a comparison task, the number 1 is always small, the number 2 is small most of the time and so on (Verguts & Fias, 2004; Verguts et al., 2005). The DSS can also handle the size effect: Its source is the everyday frequency of the numbers according to which more frequent numbers like 1 or 2 will be easier to process than rarer numbers such as 8 or 9 (Dehaene & Mehler, 1992). Without further elaboration, the DSS can accommodate a wide range of other numerical and non-numerical effects such as the distance effect in Vigliocco et al.'s (2002) study, the distance effect for letters and artificial symbols (or, in fact, any distance effect) (Jou, 2003; Razpurker-Apfeld & Koriat, 2006; Tzelgov et al., 2000; Van Opstal et al., 2008) the SNARC effect (Dehaene, Bossini, & Giraux, 1993; Krajcsi, Lengyel, & Laczkó, 2018; Leth-Steensen, Lucas, & Petrusic, 2011; Patro, Nuerk, Cress, & Haman, 2014; Proctor & Cho, 2006), the size-congruity effect (Henik & Tzelgov, 1982), the priming distance effect (Koechlin, Naccache, Block, & Dehaene, 1999; Reynvoet & Brysbaert, 1999), thus making it a comprehensive model of numerical cognition. The delta-rule connectionist model and the primitives account are both feasible implementations of the DSS.

Differentiating between the ANS and the DSS.

At this point, we have two possible accounts about what the mechanism is behind symbolic numerical cognition that can explain the comparison distance and size effects generated by symbolic numbers. This poses the question: How do we differentiate between them? The first step was choosing the appropriate paradigm. The two effects are traditionally obtained in the number comparison task, so it made sense to investigate the putative sources of the two effects and thus draw conclusions about the underlying mechanism through that task. The number comparison task is probably the most widely used experimental paradigm in numerical cognition. In its most common version two numbers are compared by choosing the numerically larger of the two.

Quantifying the two models and comparing them directly to test which one better describes the performance in a number comparison task was the next step. In the case of the ANS, the literature offers several options for the quantitative description of performance (Dehaene, 2007; Kingdom & Prins, 2010) that are elaborated upon in Thesis Study 1. The DSS is, at present, underspecified, and as such cannot offer a specific proposition as to what exact mathematical formula could describe the distance and the size effects and their relationship – even if we know their possible sources, there is no precedent in the literature that can present a quantitative description. Thus, our quantitative proposal is unavoidably speculative. Nevertheless, the DSS account does impose a few constraints. First of all, independent of what its source might be, a distance effect is expected to appear, and can be simply quantified as the absolute distance between the two numbers to be compared. Second, the frequency of the numbers have been empirically measured in different language corpora (Dehaene & Mehler, 1992), and the frequency term for each number can be approximately quantified as the number on the power of -1. Thus, the size effect can be quantified as the sum of the frequencies of the two numbers. Then the two can be added up as the simplest mathematical solution. This allowed for constructing testable predictions for both the ANS and the DSS, even though the patterns predicted by the two models were very similar. The fit to the data in the study did not offer an unequivocal resolution, but was a good starting point to find other ways to differentiate between the two models.

As the two effects themselves are what leads to conclusions about the representation of numerosity, this means that they are feasible tools for differentiating between the ANS and the DSS. The best way to do that was that since there are predictions about their sources, we could design experimental manipulations based on those predictions, and see whether the results support either model. In the case of the size effect, we manipulated the frequency with which the numbers were presented (Thesis Studies 1 and 4). In the case of the distance effect, we manipulated the associations between the numbers and the “small-large” properties by presenting only a partial number set (numbers 1, 2, 3, 7, 8, and 9) in the comparison task (Thesis Studies 2 and 3). According to the ANS the two effects stem from the same source, the ratio between the numbers, and should not be affected by these manipulations. Furthermore, our designs manipulated the possible sources independently, thus allowing us to explore whether the effects change independently of each other, which would strengthen the claim that there is no common source.

Indo-Arabic digits supposedly have established everyday frequency and values or stable connections with the "small-large" properties because of their everyday use. For that reason, our experiments at first investigated the changes in the distance and size effects in new, artificial numbers (Thesis Study 1 for the size effect and Thesis Study 2 for the distance effect). The new numbers do not have previously established frequency or associations, although they could take on those of the Indo-Arabic notation in which case the results should be very similar to those for Indo-Arabic numbers. One crucial point here was to determine whether the new symbols bear the same meaning as the Indo-Arabic numbers. An additional priming task in Thesis Study 1 was introduced to ensure that the new digits are not a series of symbols independent of their intended values, but they can be considered as a notation for the respective numbers. The priming distance effect (PDE) is observed when the closer in value the prime is to the target, the faster the response to the target is, e.g., the response to the target 6 is faster when the prime is 7 than when the prime is 9. It is considered to be a sign of the semantic relationship between the symbols or the overlap of their representations (Van Opstal et al., 2008), i.e., the symbols with the same value attached to them activate the same representation. The results from the two studies showed that the manipulations had the expected effect on artificial numbers, and that prompted the question whether what is considered as stable in the Indo-Arabic numbers really is that. The exact same manipulations were conducted with Indo-Arabic numbers, first for the distance effect (Thesis Study 3), then for the size effect (Thesis Study 4). Table 1 summarizes which effect was investigated in each study. The aims of the studies are summarized in the following chapter.

Table 1. Summary of which effect was studied in each study and which notation was used.

	Distance effect	Size effect
New symbols	Thesis Study 2	Thesis Study 1
Indo-Arabic digits	Thesis Study 3	Thesis Study 4

Aims

The overall aim of the studies was to investigate the source or sources of the distance effect and the size effect in symbolic numerical cognition. In the presented Thesis Studies the aims were as follows:

1. examine whether a different model (DSS) is a better description for the data obtained in the symbolic number comparison task than the ANS (Thesis Study 1, Experiment 1);
2. examine frequency as a possible source of the size effect by testing whether it can be induced by manipulating the frequency of presentation of the numbers when recently learned artificial numbers are used for which there is no prior experience (Thesis Study 1, Experiment 2 and 3);
3. examine the associations between the numbers and the "small-large" properties as a possible source of the distance effect by manipulating those associations in a new, artificial number sequence (Thesis Study 2);
4. examine whether the associations between numbers and the "small-large" properties can be modified in Indo-Arabic numbers within a session of the comparison task, i.e., seek further confirmation for the distance effect being association-based (Thesis Study 3);
5. examine whether the size effect shows similar flexibility in Indo-Arabic numbers by manipulating the frequency of presentation of the numbers within a session, i.e. further confirmation for frequency being the source of the size effect (Thesis Study 3 and Thesis Study 4);
6. examine whether the distance and the size effects change independently of each other (all Thesis Studies);
7. more generally, an aim present in all reported studies, contrast the two proposed models of numerical cognition, the ANS and the DSS, in symbolic numbers. Here, the sources of the numerical distance and size effects are examined for being consistent with either account, and conclusions about the two accounts will be drawn based on that, but any further investigation of the two models is beyond the scope of the thesis.

Brief Summary of the Studies[2]

In Thesis Study 1 we started with a direct comparison of two possible models of symbolic numerical cognition – the ANS and the DSS – on behavioral data from the number comparison task. Each of the two models suggests different sources for the distance and the size effects, thus if any of the two gave a better explanation of the data, then it can be claimed that that explanation is the mechanism behind symbolic numerical cognition. Several possible descriptions of the ANS model found in the literature and possible descriptions of the DSS model are tested. The two models, however, proved to be very similar. A possible method to investigate the two models was to introduce new, artificial symbols instead of the overlearned Indo-Arabic numbers. One property that can be manipulated in new symbols is the frequency with which they are presented to the participants. New symbols do not have a previously established frequency, whereas the frequency of Indo-Arabic numbers has been investigated in language corpora (Dehaene & Mehler, 1992) and has been shown to be similar in several languages. Participants compared numbers (new symbols) presented with either uniform or Indo-Arabic-like (everyday) frequency. The results showed that it was the frequency that induced a size effect.

We then turned to the distance effect in Thesis Study 2. Using new symbols, only the numbers 1, 2, 3, 7, 8, and 9 were taught to the participants and then presented in the comparison task. Thus, 3 and 7 were neighbors in the sequence, but not neighbors by value. In the number comparison task this manipulation changes the associations between the numbers and the "small-large" properties. If the source of the distance effect is the value, i.e., the meaning of the number as the ANS predicts, then performance for number pairs from the opposite sides of the gap should not change with the removal of the numbers in the middle. The DSS account allows for two predictions: 1) the new numbers acquire the meaning and/or the associations of the Indo-Arabic numbers, and there will be no change in performance, or 2) the new numbers form their

---

2 The text of the first two Thesis Studies is the final, published version (both are in an Open Access journal). The text of the last two Thesis Studies is the accepted version of the manuscript, but not the published version as their copyright has been transferred to the publisher. The figures and tables are embedded in the text, and have been renumbered (as compared to the published papers), so that they are consequently numbered within the thesis. All figures have been recreated to avoid copyright issues. The only changes made to the text, figures, and tables are formatting changes. All references have been moved to a separate section at the end. All references have been updated to their latest status.

own associations with the “small-large” properties, and, for example, performance for the 3-7 number pair would be the same as if the numerical distance between them was 1 instead of 4. The results of Thesis Study 2 supported the second possibility, thus establishing associations as the source of the distance effect, and supporting the DSS model. A further observation in line with Thesis Study 1 was that with new symbols and a uniform frequency of the numbers the size effect did not appear.

The result of Thesis Study 2 raised the following question – are the associations between numbers and the “small-large” properties stable or can they be modified within one session of the comparison task? We formerly supposed that in Indo-Arabic numbers the associations are stable and correlate highly with the value of the numbers, but this has not been previously investigated. In Thesis Study 3 we used the same methods as in Thesis Study 2, but with Indo-Arabic numbers as stimuli. The change in associations was observable already in the beginning of the session, and remained stable over the course of the experiment. The distance effect was again shown to be association-based and in addition to that, flexible.

An observation in Thesis Study 3 was that the size effect appeared and remained stable despite the uniform frequency of the numbers. However, the distance effect was shown to be flexible in Indo-Arabic numbers, and the presence of the size effect could be manipulated in new symbols. Thesis Study 4 examined whether this effect can be altered in Indo-Arabic numbers whose frequency in everyday language is established. Participants compared numbers in three conditions: in the first the numbers had uniform frequency, in the second everyday frequency, and in the third reversed everyday frequency. According to the results the size effect decreases, but does not disappear. This suggests two components for the size effect: a flexible one which can be explained by the DSS, and a stable source which can be in line with both the ANS and the DSS models.

General Methods

This section describes the methodology of the Thesis Studies in general.

Participants.

Participants were students participating for partial credit. The aim was to have 15-20 participants per group in an experiment as preliminary research in the laboratory had established that this number is sufficient for obtaining reliable distance and size effects.

Stimuli.

The stimuli were either Indo-Arabic numbers or artificial numbers (Ⴚ, Ω, Ꭷ, Ծ, Ճ, Ł, Ɵ, Ð, И, Я, Ч, Ħ, Œ, ƌ, ƛ, ʕ, Ʊ, Ʋ) that the participants learned. The symbols were chosen from non-Latin alphabets as it has been demonstrated that letters from alphabets already in use resemble configurations from our natural environment, and are thus more appropriate to use as substitutes of already familiar symbols (Changizi & Shimojo, 2005). The symbols had similar width and height. The aim of each study determined whether the whole set of digits from 1 to 9 or only a partial set (1 to 3 and 7 to 9) was used. The artificial symbols were chosen randomly from the list of twenty symbols seen here, and then randomized for each participant. This way no participant was given the same list of symbols for the same numbers.

The initial choice of new symbols was made to avoid the inevitable interference of the already established frequency and associations of Indo-Arabic numbers from a lifetime of everyday practice. New symbols learned as numbers have been shown behaviorally to exhibit the same effects as Indo-Arabic numbers, such as the distance effect, the number congruence effect, the priming distance effect in preliminary studies conducted in our laboratory. After the results in Thesis Studies 1 and 2 pointed to the frequency and the associations with the “small-large” properties being the sources of the size effect and the distance effect respectively, we questioned our presumption. In other words, it is possible that the associations are flexible, depending on the situational context, i.e., they can be changed within the comparison task, and thus alter the distance effect. The same could be said about the frequency – biasing the frequency of the presented numbers in the opposite direction of their everyday frequency may cause the size effect to disappear. Thus, we continued our explorations with Indo-Arabic numbers as well.

Tasks.

Three types of tasks were employed in the Thesis Studies. In the case of experiments with artificial symbols, the participants were presented with a list of paired Indo-Arabic numbers and new symbols to learn. They could look at the list for as long as necessary, and then start the learning task during which an Indo-Arabic number and a new symbol were displayed on the screen. The participants had to decide whether the pair denoted the same number by pressing a key for “same” or “different”. In half of the trials (5 trials for each number per block) the correct response was “same”. The learning task had five blocks, and continued until the participant reached a lower-than-five-percent error rate in a block or did not reach the threshold until the end of the last block.

The second task was the number comparison task in which the participant saw two numbers (both either Indo-Arabic digits or new symbols) and had to decide which one is the larger numerically by pressing a key. All pairs excluding ties from the relevant set of stimuli were presented (but see the Methods section of each experiment for details). This was the main task for all presented studies. Here is where the experimental manipulation occurred – the presentation frequency of the numbers, the associations of the numbers, the length of the experiment (number of trials) were manipulated. The latter manipulation was important – for Indo-Arabic numbers we wanted to explore whether a change in the distance and size effects would happen gradually or immediately, if at all.

The third task, a priming task, was used only in Thesis Study 1 to determine whether the new symbols were linked semantically to the Indo-Arabic numbers (Koechlin et al., 1999; Reynvoet & Brysbaert, 1999). A prime (new symbol) was shown to the participants, and was then followed by a target Indo-Arabic number. For both prime and target the participants decided whether the number was smaller or larger than 5 with a key press. The prime and the target remained visible until response with a 200 ms interval between the response for the prime and the target and a 2000 ms interval between the response for the target and the next prime. The appearance of the priming distance effect (i.e., the error rate and rthe eaction time for the target depended on the numerical distance between the prime and the target with smaller numerical distance resulting in faster responses and fewer errors) shows that the meaning of the new symbol primed the meaning of the Indo-Arabic number.

In the case of new symbols, auditory feedback was given for correct and incorrect responses in all tasks.

Presentation of the stimuli and measurement of the responses were managed by the PsychoPy software (Peirce, 2007). Data processing and statistical analyses were performed in LibreOffice (LibreOffice, 2018), CogStat (Krajcsi, 2018), G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007), and SPSS (IBM Corp., 2012, only for the repeated measures analysis of variance calculations).

Procedure.

The participants were recruited via a freely elective course in which participating in experiments was worth partial credit towards the final grade. The experiment was announced on the course's forum which automatically sent an e-mail to the students taking the course. The participants signed for the experiment by e-mail for the given dates. Before the experiment they were given a brief description and signed an informed consent. The experiment commenced and depending on the study, lasted between 30 and 90 minutes. The participants could ask questions at any time before, during, and after the experiment. In Thesis Study 1 and the first experiment in Thesis Study 2 the collection of data happened individually, and in Thesis Studies 3 and 4 and in the replication experiment in Thesis Study 2 the data collection happened in groups due to a change in the laboratory's data collecting procedure.

Once the participants signed the informed consent, the experimenter recorded their demographic data – gender, age, handedness, vision status, university program, any additional notes. A code was assigned to each participant so that the data could not be traced back to the participant. Then the experimenter started the experiment. If the study contained only the comparison task, then the participant just completed the task. In case there was a learning task at the beginning, with the end of that task the experimenter started the comparison task. The priming task was also started separately, always as the last one.

Measurement.

Error rates and reaction times.

For all tasks errors and reaction times were recorded for each trial. Error rates were calculated for each participant, and then used in the analysis. In Thesis Study 1 median reaction times for the correct responses were computed for each participant, whereas in the other three studies all trials with reaction time over 2000 ms were removed, and then mean reaction times for the correct responses were calculated for each participant. The switch was mostly due to one additional measurement, the drift rate (see the next section), that we included in the analysis. Drift rate is derived from the proportion of correct responses, reaction time variance, and number of correct trials, thus using mean reaction time seemed the more sensible choice.

Drift rate (Diffusion model analysis).

Drift rate is a part of the increasingly popular diffusion model analysis, and is assumed to provide a more sensitive measure of performance (Ratcliff & McKoon, 2008; Smith & Ratcliff, 2004). In that model evidence is accumulated gradually from perceptual and other systems until a sufficient amount of evidence becomes available for a decision to be made. Drift rate represents the quality of information upon which the evidence is built, and while error rates and reaction times adequately capture performance on a task, drift rate is more directly related to the background mechanisms of performance. Furthermore, drift rates can be recovered based on observed error rate and reaction time parameters (Ratcliff & Tuerlinckx, 2002; Wagenmakers, Van Der Maas, & Grasman, 2007). In the presented studies, the EZ-diffusion model (Wagenmakers et al., 2007) is applied, a simplified version of the diffusion model which still allows for the recovery of drift rates in the case of sparse data from a relatively small number of parameters. For edge correction we used the half-trial solution (see the exact details about edge correction in Wagenmakers et al., 2007). The scaling within-trials variability of drift rate was set to 0.1 in line with the tradition of the diffusion analysis literature.

General methods for analysis.

Analysis of the number comparison task.

Traditional methods for analyzing the distance and the size effects and the full stimulus space.

Traditionally, when analyzing the data from the comparison task, the results are collapsed across distances and sizes (e.g., performance for the 7-8 and the 4-5 pairs is combined under numerical distance 1, but is added to numerical size 15 and 9 respectively). This leads to loss of information, possible artifacts, cells with unequal weights. In our experiments the full stimulus space was used instead, and the classical analysis was added only as illustration. The full stimulus space is a grid (matrix) in which the rows denote one of the members of the number pair, the columns denote the other, and the cells denote performance (Figure 2). The full stimulus space is a data-driven approach to analysis, which helps ensure that the effects are not due to artifacts and important information is not omitted. It allows for more precise regression – any predictor can be used, predictors can be combined, and the cells have equal weights. It displays systematic patterns which may otherwise remain unobserved, and helps to communicate the results in a more visually accessible manner Most importantly, evidence from the full stimulus space is more reliable and convincing as it is possible to observe all systematic patterns, which is extremely useful when we have to be especially careful that the effects are not due to artifacts. This method is applied in all experiments.

The full stimulus space

Number 1

Number 2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1		0.3	0.2	0.1	0.1	0.1	0.1	0.1	0.1
2	0.3		0.5	0.3	0.2	0.2	0.1	0.1	0.1
3	0.2	0.5		0.6	0.4	0.3	0.2	0.2	0.2
4	0.1	0.3	0.6		0.7	0.5	0.4	0.3	0.3
5	0.1	0.2	0.4	0.7		0.8	0.5	0.4	0.4
6	0.1	0.2	0.3	0.5	0.8		0.8	0.6	0.5
7	0.1	0.1	0.2	0.4	0.5	0.8		0.9	0.7
8	0.1	0.1	0.2	0.3	0.4	0.6	0.9		1.0
9	0.1	0.1	0.2	0.3	0.4	0.5	0.7	1.0	

Figure 2. An illustration of the full stimulus space. Columns indicate one number of the pair to be compared, rows indicate the other number, and cells show expected performance. Darker shade indicates worse performance. The distance effect can be observed as better performance from the main diagonal towards the top-right and bottom-left corners. The size effect is worse performance along the main diagonal from the top-left toward the bottom-right corner. This figure is a copy of the left panel of Figure 4B (Thesis Study 1), and the values were calculated as a × log(large/distance) + b, where a is set to 1 and is b is set to 0.

Model comparison.

Linear regression model fitting was used in the studies to decide which of the predictions in the respective Thesis Study is a better fit for the data. Regressors were defined for the full stimulus space and fitted to the data. Goodness of fit (R^2) was computed for each model for the averaged results (group level), and then for each participant (individual level). The latter allowed for a statistical comparison of the models with a non-parametric statistical test in which the R^2s were considered ordinal variables. In all cases the linear fit was conducted with the least square method.

Presence of the effects.

When investigating the presence of the effects, the slopes of the distance, size, and priming distance effects were calculated by fitting a linear regression model to the data, and then the deviation of the slopes (the beta weights) from 0 was tested for each effect (i.e., whether the regressor significantly contributed to the explanation of the variance which meant that the effect was present).

One point of consideration is how the regressors are expressed. The distance in symbolic numbers is traditionally computed as the absolute difference between the numbers to be compared. However, it is likely that a logarithm of that difference is a more appropriate expression as reaction time cannot decrease indefinitely. In the reported studies either both linear distance and its logarithm or only the logarithm version are used as regressors. The size effect is usually expressed as the sum of the numbers to be compared and is used as such in the studies. Presently, a more appropriate calculation method is not available.

An additional effect which has to be taken into consideration is the end effect (Pinhas & Tzelgov, 2012). For the instruction "choose the larger" participants tend to be much faster for the cells containing the largest number of the set. This is a non-linear distortion of the full stimulus space. Each of the studies checks the influence of the end effect, and if necessary, solves for its presence by either removing the concerned cells or by including the effect as a regressor in a multiple linear regression[3].

Meta-analysis.

In some of our studies replication was requested by the reviewers as there were no earlier studies that conducted experiments and analysis similar to ours. This allowed for running a mini meta-analysis on the data (e.g., Goh, Hall, & Rosenthal, 2016; Maner, 2014).

Calculation of the average values.

For the comparison task, average error rates were calculated first, and participants with higher than mean+2 standard deviations error rate were excluded from further analysis. Average error rates, reaction times, and drift rates were calculated for each participant for the full stimulus space overall and, for blocked sessions, separately for each block. In the case of model comparison the goodness of fit (R^2), and for the

3 The changes in the utilized methods between the studies also reflect to an extent the evolution of the methods used in the laboratory.

presence of the effects the deviation of their slopes from 0 were investigated as described above.

Analysis of the learning and the priming tasks.

Learning task.

The analysis of the learning task consisted only of calculating average error rate and examining which of the participants did not meet the minimum requirement of error rate below 5% in their last block. Participants who did not fulfill the requirement were excluded from further analysis.

Priming task.

Average error rates and reaction times were calculated for each participant and for each distance between a prime and a target (the absolute difference of the two stimuli). Only trials requiring the same response (smaller or larger than 5) for prime and target were included in the analysis to avoid the effect of congruence. Zero distance pairs (same prime and target) were removed as for these pairs there was additional training in the learning task, and the slope of the change along the distance was tested for being significantly different from 0. For the priming distance effect the experimental data was collapsed according to distance. As the effects size of the priming distance effect proved to be rather small, a meta-analysis was used on the data from the experiment and data from earlier studies to investigate whether it was present as well as the necessary size of the sample for sufficient statistical power.

To sum up, the data were analyzed with both traditional and novel (to the field) methods. Thus, the results have consequences not only for the comparison of the ANS and the DSS account, but also for the methods used in numerical cognition.

Summary

The source of the numerical distance and size effects in a number comparison task is examined in four studies. There are two possible accounts. The ANS suggests that both effects are a consequence of the representation of numerosity being an innate, analogue, noisy representation that works according to Weber's law, and their source is the ratio of the numbers. According to the DSS the representation of numerosity is similar to the mental lexicon or a semantic network. The distance effect is rooted either in the connections between the nodes or their associations to the "small-large" properties, and the size effect is a result of the everyday frequency of the numbers. To differentiate between the two accounts, we manipulated the associations of the numbers and their frequency in the number comparison task in new, artificial symbols and in the Indo-Arabic notation, and then tested the prediction of which of the two accounts is a better of fit for the experimental data and how the effects are affected by the manipulation.

Thesis Study 1
The Source of the Symbolic Numerical Distance and Size Effects

Attila Krajcsi[1], Gábor Lengyel[2] and Petia Kojouharova[1,3]
[1]Department of Cognitive Psychology, Institute of Psychology, Eötvös Loránd University, Budapest, Hungary
[2]Department of Cognitive Science, Central European University, Budapest, Hungary
[3]Doctoral School of Psychology, Eötvös Loránd University, Budapest, Hungary

Article published online in *Frontiers in Psychology* on 21st November 2016
DOI: https://doi.org/10.3389/fpsyg.2016.01795

Human number understanding is thought to rely on the analog number system (ANS), working according to Weber's law. We propose an alternative account, suggesting that symbolic mathematical knowledge is based on a discrete semantic system (DSS), a representation that stores values in a semantic network, similar to the mental lexicon or to a conceptual network. Here, focusing on the phenomena of numerical distance and size effects in comparison tasks, first we discuss how a DSS model could explain these numerical effects. Second, we demonstrate that the DSS model can give quantitatively as appropriate a description of the effects as the ANS model. Finally, we show that symbolic numerical size effect is mainly influenced by the frequency of the symbols, and not by the ratios of their values. This last result suggests that numerical distance and size effects cannot be caused by the ANS, while the DSS model might be the alternative approach that can explain the frequency-based size effect.

An Alternative to the Analog Number System

According to the current models understanding numbers is supported by an evolutionary ancient representation shared by many species (Dehaene, Dehaene-Lambertz, & Cohen, 1998; Gallistel & Gelman, 2000; Hauser & Spelke, 2004), the analog number system (ANS). One defining feature of the ANS is that it works similarly to some perceptual representations in which the ratio of the stimuli's intensity determines the performance (Weber's law) (Cantlon et al., 2009; Moyer & Landauer, 1967; Walsh, 2003). Two critical phenomena supporting the ratio based performance are the distance and the size effects: when two numbers are compared, the comparison is slower and more error prone when the distance between the two values is smaller (distance effect) or when the two numbers are larger (size effect), (Moyer & Landauer, 1967) (Figures 3 and 4). Thus, in the literature, the numerical distance and size effects are considered to be the sign of an analog noisy numerical processing system working according to Weber's law. The distance and the size effects are observable both in non-symbolic and symbolic number processing, reflecting that the same type of system processes numerical information, independent of the number notations (Dehaene, 1992; Eger, Sterzer, Russ, Giraud, & Kleinschmidt, 2003).

However, the distance and size effects in symbolic comparison can also be explained by a different representation. Quite intuitively, one might think that symbolic and abstract mathematical concepts, like numbers could be handled by a discrete semantic system (DSS), similar to conceptual networks or to the mental lexicon, i.e., representations that process symbolic and abstract concepts. In this DSS model, numbers are stored in a network of nodes, and the strength of their connections is proportional to the strength of their semantic relations. We propose that this DSS account could be responsible for symbolic number processing; whereas non-symbolic number processing is still supported by the ANS (see some additional details about the relation of the two models below). The main aim of the present study is to investigate the feasibility of the DSS model as a comprehensive explanation of the symbolic numerical effects, and to contrast it with the ANS model.

DSS explanation for the distance and size effects.

How can a DSS explain the symbolic numerical distance and size effects? (1) Regarding the distance effect, the strength of the connections between the nodes can produce an effect which is proportional to their strength, and since in a network storing numbers the strength of the connections is proportional to the numerical values and numerical distance, this system could produce a numerical distance effect. In fact, a similar semantic distance effect was shown in a picture naming task (Vigliocco et al., 2002): Naming time slowed down when the picture of the previous trial was semantically related to the present picture, and a small semantic distance between the previous and the actual word caused stronger effect than a large semantic distance, similar to the numerical distance effect[4]. This semantic distance effect cannot be the result of a continuous representation similar to the ANS, because the stimuli were categorical (e.g., finger, car, smile, etc.)[5]. Thus, a discrete representation potentially can produce a numerical distance effect. Several mechanisms can be imagined how a numerical distance effect is generated. One can imagine that the semantic distance information, that can be revealed in a semantic priming, could generate a distance effect. Alternatively, it is possible that the strength of the association between the numbers and the large–small categories create the numerical distance effect (Verguts & Fias, 2004; Verguts et al., 2005). Here, we do not want to specify the exact mechanism behind the numerical distance effect, but only propose that several possible mechanisms are already available in the literature. (2) Turning to the size effect, this effect also could be generated by a DSS. It is known that smaller numbers are more frequent than larger numbers, and the frequency of a number is proportional to the power of its value (Dehaene & Mehler, 1992). Since the numbers observed more frequently could be processed faster, the size effect could result from this frequency pattern[6]. Thus, the DSS

4 Comparison distance effect (e.g., which of two numbers is larger) and priming distance effect (whether previous stimulus influences the actual stimulus processing based on the distance of the two stimuli) are known to be two different mechanisms (Reynvoet, De Smedt, & Van den Bussche, 2009; Verguts et al., 2005). While we want to find a DSS explanation for the comparison distance effect, the cited semantic distance effect is more similar to a priming distance effect. Importantly, we are not stating that these two effects are the same, but we suggest that a distance-based effect is possible in a DSS, independent of the exact mechanism behind that effect.

5 A similar proposal is that the numerical distance effect might emerge from the order property of numbers, and a distance effect can be observed not only in numbers, but also in non-numerical orders, e.g., days or letters (Potts, 1972; Verguts & Van Opstal, 2014). However, (a) it might be possible that in those examples the non-numerical orders are transformed to the numerical representation, which is not possible for the categorical words in the cited picture naming task (Vigliocco et al., 2002), and (b) the DSS model has less strict constrains, i.e., no order structure is presupposed, but a more general series of associations is sufficient to explain the distance effect.

6 Frequency is essential in other numerical tasks to produce size effect (Zbrodoff & Logan, 2005), and the role of frequency in size effect was also proposed in other alternative models of number

model can also explain the appearance of distance and the size effects (Figure 3).

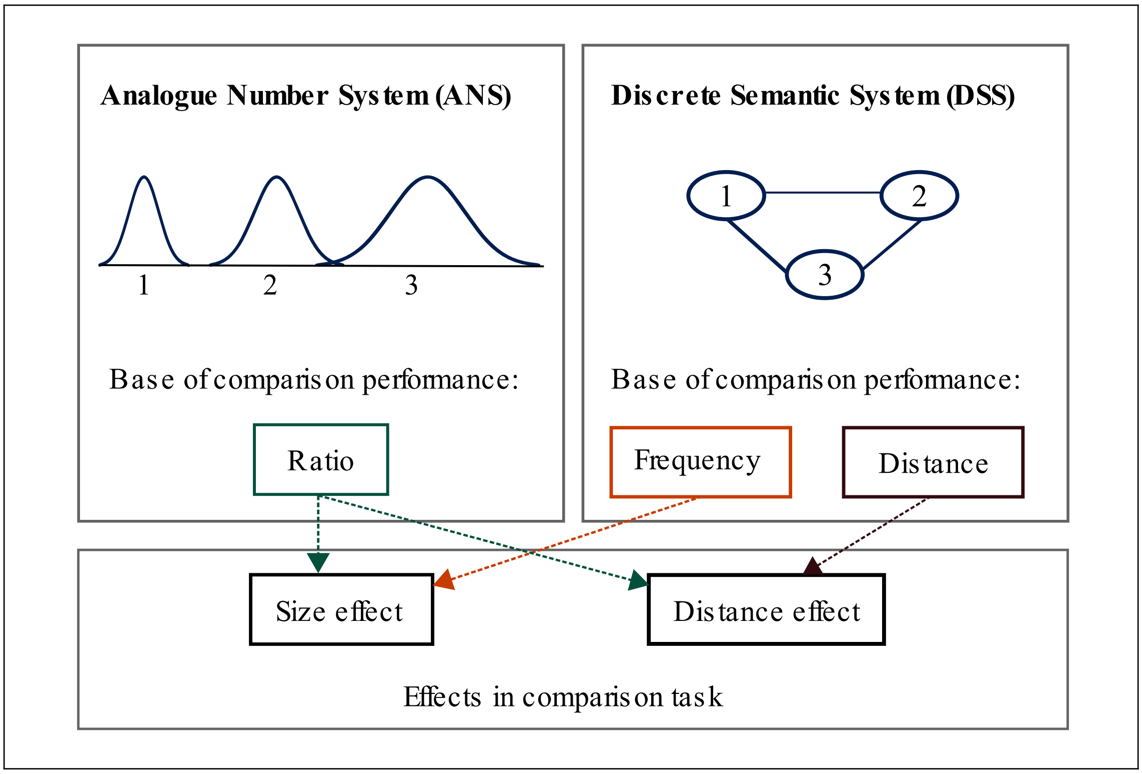


Figure 3. The sources of the distance and size effects according to the two models.

DSS explanation for other numerical effects.

Whereas in the present work we focus on the DSS explanation of the distance and size effects, the DSS explanation can be readily extended to other effects, too, and it can be a comprehensive model of symbolic number processing. The following details can demonstrate that despite its radical difference from the ANS model, DSS might be a viable option to explain symbolic numerical phenomena. Many of these explanations have already been proposed in the literature, although these explanations usually focused on single specific phenomena, and they did not offer a comprehensive model.

Several interference effects can be explained in the DSS framework. For example, the SNARC effect (interference between numerical value and response location in a task) was originally interpreted as the interference of the ANS's spatial property and the response locations (Dehaene et al., 1993), however, it is also possible that the effect is the result of the interference of the left-right and large-small nodes in a semantic network similar to the DSS (Krajcsi, Lengyel, & Laczkó, 2018; Leth-Steensen et al., 2011; Patro et al., 2014; Proctor & Cho, 2006). Similarly, while the size

comparison (Verguts & Van Opstal, 2014).

congruency effect (Stroop-like interference between the numerical value and the physical size of symbols; Henik & Tzelgov, 1982) can be thought of as an interference between the ANS and a representationally similar analog size representation, it can also be thought of as an interference between the many-few and the physically large-physically small nodes.

While there are many empirical and theoretical works in the literature that support the ANS model, in fact there are only a handful effects that are cited to support the ANS model, and we propose that most of these effects (in fact to our knowledge all of them at the moment) can also be explained by the DSS. While mostly it would not be too difficult to find DSS explanations for different phenomena, in the present work we only focus on the numerical distance and size effects in comparison tasks.

Different representations for symbolic and non-symbolic numbers.

As it was mentioned above, the DSS model can only account for symbolic number processing. Clearly, there are cases when the DSS cannot handle numerical information, for example, when the symbolic mental tools are not available, like in the case of infants (Feigenson et al., 2004), animals (Hauser & Spelke, 2004), or adults living in a culture without number words (Gordon, 2004; Pica et al., 2004), therefore, the ANS seems to be a sensible model to explain these non-symbolic phenomena. It also seems reasonable that because of their representational structure, the two systems could be specialized for different forms of numbers: The DSS could be responsible for the precise and symbolic numbers, while the ANS could process the approximate non-symbolic stimuli.

This idea of different representations for symbolic and non-symbolic numbers is supported by the increasing number of findings in the literature, suggesting that symbolic and non-symbolic number processing is supported by different representations. For example, it has been shown that performance of the symbolic and non-symbolic number comparison tasks do not correlate in children (Holloway & Ansari, 2009; Sasanguie et al., 2014), and in an fMRI study the size of the symbolic and non-symbolic number activations did not correlate (Lyons, Ansari, & Beilock, 2015). As another example, whereas former studies found common brain areas activated by both symbolic and non-symbolic stimuli (Eger et al., 2003; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004), later works with more sensitive methods found only notation-dependent activations (Bulthé et al., 2014; Bulthé et al., 2015; Damarla & Just, 2013).

According to an extensive meta-analysis, although it was repeatedly found that simple number comparison task (the supposed sensitivity of the ANS) correlates with mathematical achievement, it seems that non-symbolic comparison correlates much less with math achievement, than symbolic comparison (Schneider et al., 2017). In another example, Noël and Rousselle (2011) found that whereas older than 9- or 10-year-old children with developmental dyscalculia (DD) perform worse in both symbolic and non-symbolic tasks than the typically developing children; younger children with DD perform worse than control children only in the symbolic tasks, but not in the non-symbolic tasks. The authors concluded that the deficit in DD can be explained in the terms of two different representations: The deficit is more strongly related to the symbolic number processing, and the impaired non-symbolic performance is only the consequence of the symbolic processing problems. See a more extensive review of similar findings in Leibovich and Ansari (2016). All of these findings are in line with the present proposal, suggesting that symbolic and non-symbolic numbers are processed by different systems.

Related models for symbolic number processing.

There are former models in the literature that are potential alternatives to the ANS model, and some of those models can be fitted into a DSS framework, or they could be considered as implementations of more specific aspects of the DSS account.

Verguts et al. (2005) and Verguts and Van Opstal (2014) proposed a connectionist model describing several phenomena of number processing and more generally several phenomena of ordinal information processing. According to their simulations and experiments, this model offers a superior description of number naming, parity judgment and number comparison than the ANS model, and their model can also explain non-numerical order processing phenomena. Their model includes a hidden layer representing the values of the numbers in a place-code with a fixed width of noise. This means that the nodes of the hidden layer represent numbers on a linear scale, and a number most strongly activates the node mainly representing that number, but additional activation also can be found in the neighboring nodes. The distance these additional activations can reach to do not depend on the source number, i.e., the noise has a fixed width. Although the authors suggest that this model implements an analog representation, it contradicts the ANS model, because on a linear inner scale the size of the noise is not proportional to the size of the number, and relatedly it could not generate ratio-based performance. In line with this representational issue, the model in itself cannot produce a size effect, and an uneven frequency of numbers should be introduced to generate the numerical size effect (Verguts & Fias, 2004; Verguts et al., 2005), questioning whether this model can be seen as an ANS-like model. However, we propose here that the model can be interpreted as a discrete symbolic representation: Activation in the neighboring nodes is not the noise of that representation but it is a spreading activation in the hidden layer. With this alternative interpretation the model can be seen as a specific implementation of the discrete symbolic system when stimuli are arranged as an ordered list. Note that in their model the comparison distance effect is not explained by the spreading activation, but by the connection weights between the value nodes and the response nodes (Verguts et al., 2005; Verguts & Van Opstal, 2014). This model as a potential DSS implementation can give a more precise description for a whole range of phenomena, the ANS model could not account for, thus, strengthening the DSS explanation of symbolic number processing.

Tracking a different line, Henik and Tzelgov (1982) investigated automatic processing of numbers with the size congruency effect (interference between the physical size and numerical value properties of the stimuli). Based on their results they suggested that some basic elements (primitives) are stored in the long term memory, e.g., integers from 1 to 9 and the number 0 (Pinhas & Tzelgov, 2012), while other numbers are not stored as basic elements, e.g., negative numbers and ratios (Kallai & Tzelgov, 2009; Tzelgov, Ganor-Stern, & Maymon-Schreiber, 2009). The basic elements or primitives can be considered as the nodes of the DSS: These basic elements could be the values that are stored in the nodes of the network, while other numbers are the combination of the primitives, somewhat similar to the relation of words and sentences. Also, the size congruency effect can be used as a method to find whether a number is stored as a unit in the DSS.

Possible quantitative descriptions of symbolic comparison performance in the DSS model.

While the DSS model can explain why the numerical distance and size effects appear in a comparison task, the ANS model not only suggests that there should be numerical distance and size effects, but it offers a quantitative description for the performance. For example, Moyer and Landauer (1967) proposed that the reaction time of a comparison task is proportional to the following function:

$K \times \log\left(large_{number} / \left(large_{number} - small_{number}\right)\right)$ (See Dehaene, 2007 for a more detailed description of the ANS predictions for behavioral numerical decisions.)

One of the next challenges for the DSS model is to find a quantitative description similar to the ANS model. As in the ANS model where the details of the model were borrowed from psychophysics models, we borrow the details of the DSS model from psycholinguistics and semantic network models. Unfortunately, whereas in many cases the psychophysics models offer quantitative descriptions of the performance (Dehaene, 2007; Kingdom & Prins, 2010), the bases of the DSS model do not have consensual quantitative descriptions. Additionally, our description does not build upon a detailed working model with specific mechanisms (e.g., as it was mentioned, there could be different candidates that could generate the distance effect), but a functional description of these potential effects are given here. Thus, our quantitative proposal is unavoidably speculative, although there are some constrains we can build upon. First, one term of this quantitative description should depend on the distance between the two values. Second, another term should depend on the frequencies of the values, where the frequency of the number is the power of that number (Dehaene & Mehler, 1992). Current theoretical considerations do not specify what distance and size functions should be used, how the frequency of the two numbers should be combined, and how exactly the two terms create performance, thus these details are unavoidably speculative at the moment, and future work can refine the versions offered here. However, based on these few starting points, a number of alternative versions of the DSS model can be created, and many of them display a qualitatively similar pattern of number comparison performance. One simple example is displayed on Figure 4, where, as the mathematically simplest version, the distance effect is a linear function, the frequencies of the numbers are summed up, and the distance and size components are added up. This DSS-motivated function creates a qualitatively very similar pattern to the function of the ANS model: Looking at the patterns, the two models are rather similar, also reflected in the high correlation between the two models (r = -0.89). Thus, one can create a hypothetical quantitative description based on the DSS account that seemingly can explain the comparison performance in a similar way as the ANS model[7].

7 After creating additional versions of the DSS quantitative prediction with considering the constrains described here, we found qualitatively similar patterns. See another example in the Methods section of Experiment 1.

A

Analogue Number System (ANS) model	**Discrete Semantic System (DSS) model**
RT = a × log(large/distance) + b	RT = a_1×distance + a_2×(x_1^{-1}+x_2^{-1}) + b

B

Number 1

Number 2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1		0.3	0.2	0.1	0.1	0.1	0.1	0.1	0.1
2	0.3		0.5	0.3	0.2	0.2	0.1	0.1	0.1
3	0.2	0.5		0.6	0.4	0.3	0.2	0.2	0.2
4	0.1	0.3	0.6		0.7	0.5	0.4	0.3	0.3
5	0.1	0.2	0.4	0.7		0.8	0.5	0.4	0.4
6	0.1	0.2	0.3	0.5	0.8		0.8	0.6	0.5
7	0.1	0.1	0.2	0.4	0.5	0.8		0.9	0.7
8	0.1	0.1	0.2	0.3	0.4	0.6	0.9		1.0
9	0.1	0.1	0.2	0.3	0.4	0.5	0.7	1.0	

Number 1

Number 2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1		1.9	2.1	2.5	2.8	3.2	3.5	3.9	4.3
2	1.9		1.2	1.6	1.9	2.3	2.6	3.0	3.4
3	2.1	1.2		1.0	1.3	1.7	2.1	2.5	2.8
4	2.5	1.6	1.0		0.9	1.2	1.6	2.0	2.4
5	2.8	1.9	1.3	0.9		0.8	1.1	1.5	1.9
6	3.2	2.3	1.7	1.2	0.8		0.7	1.1	1.5
7	3.5	2.6	2.1	1.6	1.1	0.7		0.7	1.1
8	3.9	3.0	2.5	2.0	1.5	1.1	0.7		0.6
9	4.3	3.4	2.8	2.4	1.9	1.5	1.1	0.6	

Figure 4. ***(A)*** *Reaction time (RT) function for the ANS model (based on (Crossman, 1955; Moyer & Landauer, 1967) (left) and a hypothetical RT function for the DSS model where the reaction time is proportional to a combination of the specific forms of the distance and the frequencies of the numbers (right). Notations: large: larger number; distance: distance between the two numbers;* x_1 *and* x_2*: the two numbers;* a, a_1, a_2 *and* b *are free parameters.* ***(B)*** *The prediction of the models on a full stimulus space in a number comparison task of numbers between 1 and 9. Numbers 1 and 2 are the two values to be compared. Lighter shade denotes fast responses, darker shade denotes slow responses (note that numerically the ANS function increases, and the DSS function decreases toward the high ratio, but the direction is irrelevant in the linear fit below). The distance effect can be seen as the gradual change when getting farther from the top–left bottom–right diagonal, and the size effect is seen as the gradual change from top–left to bottom–right. In the figures the parameters* a *and* a_2 *are set to 1,* a_1 *is 0.4, and parameter* b *is set to 0.*

In the first section, so far we have introduced the DSS model, an alternative to the ANS explanation of number processing, where the basic building blocks of the representation are nodes with appropriate connections. We have reasoned that the DSS framework can be a comprehensive explanation of symbolic number processing. While focusing on the comparison distance and size effects, we have demonstrated that the DSS model is capable of giving as appropriate a description of the comparison performance as the ANS model. In the following parts we turn to empirical tests. First, we investigate which model describes better an Indo-Arabic comparison task. Then, we investigate a very specific aspect of number comparison where the two models have clearly different predictions: Whether the size effect depends on the frequency of the numbers (predicted by the DSS model) or on the ratio of the numbers (predicted by the ANS model).

Experiment 1 – Goodness of the Two Quantitative Description of the Models in Indo-Arabic Comparison

After creating a quantitative description for the DSS model, we can contrast the two models, testing which model (Figure 4) fits better the empirical data in an Indo-Arabic number comparison task. Although the two models strongly correlate, and the differences between them are subtle, still, there are differences between them, and it is possible that those differences are detectable in a simple comparison task, supposing that the noise is relatively low.

Methods.

Participants.

Twenty university students participated in the study. Pilot studies with Indo-Arabic and new symbols (see also the second experiment) aiming to refine the applied paradigms revealed that the main effects to be observed can be detected reliably with a sample size of around 20. After excluding two participants because of a higher than 5% error rate, the sample included 18 participants (15 females, mean age 21.5 years, standard deviation 2.8 years). All studies reported here were carried out in accordance with the recommendations of the Department of Cognitive Psychology ethics committee with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki

Stimuli and procedure.

The participants compared Indo-Arabic number pairs. In a trial two numbers between 1 and 9 were shown until response and the participants chose the larger one. All possible number pairs including numbers between 1 and 9 were shown 10 times, excluding ties, resulting in 720 trials. Presentation of the stimuli and measurement of the responses were managed by the PsychoPy software (Peirce, 2007).

Analysis methods.

In the analysis, we contrasted the two models with analyzing the reaction times, the error rates, and the diffusion analysis drift rates. (1) Reaction time analysis was used, because response latency may be a more sensitive measurement than the error rate, and the results are comparable with many former results, including the seminal Moyer and Landauer (1967) paper. However, there is no strong consensus which function could describe the ANS model (see the applied version below). (2) Error rate analysis was chosen, because the function describing error rate performance is well established (Dehaene, 2007; Kingdom & Prins, 2010), even if the measurement is not as sensitive as the reaction time data. (3) Finally, drift rate was applied, because diffusion analysis is thought to be more sensitive than the error rate or the reaction time, although its parameter recover methods could be debated. In the recent decades, the diffusion model and related models became increasingly popular to describe simple decision processes (Ratcliff & McKoon, 2008; Smith & Ratcliff, 2004). In the diffusion model, decision is based on a gradual accumulation of evidence offered by perceptual and other systems. Decision is made when an appropriate amount of evidence is accumulated. Reaction time and error rates partly depend on the quality of the information (termed the drift rate) upon which the evidence is built. Importantly for our analysis, observed reaction time and error rate parameters can be used to recover the drift rates (Ratcliff & Tuerlinckx, 2002; Wagenmakers et al., 2007). Drift rates can be more informative than the error rate or reaction time in them, because drift rates reveal the sensitivity of the background mechanisms more directly (Wagenmakers et al., 2007).

Because different versions of the ANS models and the DSS models can be proposed, multiple versions of the models were tested, when it was necessary. For the ANS model the following functions were used in the analysis. (1) Regarding the reaction time analysis, although there are several considerations how to describe the reaction time function of continuous perceptual comparisons (Crossman, 1955; Dehaene, 2007; Welford, 1960), it is not straightforward which version should be applied to describe the ANS model (Kingdom & Prins, 2010). First, we used the version used by Moyer and Landauer (1967), displayed in Figure 4. Second, we applied the $RT \propto 1/(\log(large/small))$ function suggested by Crossman (1955), which function he finds to be more superior compared to the previous function. (2) For the error rate analysis we used the ANS model described in Dehaene (2007, equation 10), which supposes a linear scaling in the ANS,

$$p_{correct}(n_1, n_2) = \int_0^{+\infty} \frac{e^{-\frac{1}{2}\left[\frac{x-(r-1)}{w\sqrt{1+r^2}}\right]^2}}{\sqrt{-2\pi}\, w\sqrt{1+r^2}} dx$$

where n_1 and n_2 are the two numbers to be compared, r is the ratio of the larger and the smaller number, and w is the Weber ratio. (3) Regarding the drift rates, in the ANS model the stored values to be compared can be conceived as two random Gaussian variables, and the difficulty of the comparison might depend on the overlap of the two random variables: Larger overlap leads to worse performance (see the detailed mathematical description in Dehaene, 2007). It is supposed that in a comparison task the drift rate depends purely on the overlap of the two random variables (Dehaene, 2007; Palmer, Huk, & Shadlen, 2005). According to the current theories, *drift_rate = k × task_difficulty*, (Dehaene, 2007; Palmer et al., 2005), or it could also include a power term as a generalization, *drift_rate = k × task_difficulty*$^{\beta}$, although the exponent is often close to 1, thus the first, proportional model approximates the second, power model. Task difficulty is measured as stimulus strength, which is calculated with the *distance/large_number* function as suggested by Palmer et al. (2005) for psychophysics comparison. Because in an analog representation as the task becomes more difficult (i.e., the two stimuli become indistinguishable) the drift rate tends to zero, in the linear fit this means that the intercept is forced to be zero. To summarize, the *drift_rate = k × distance/large_number* function was used in the drift rate analysis fit for the ANS model.

For the DSS model, two versions were used in the analysis. First, the simple linear version was applied, as described in Figure 4. Additionally, a logarithmic version of the DSS model was also used, in which the logarithm of the two terms are used, i.e., $RT \propto \log(distance) + \log(x_1^{-1} \times x_2^{-1})$ This logarithmic version seems reasonable, because strictly speaking the distance effect cannot be linear, since that would result in negative reaction time or error performance for sufficiently large distances (even if the linear version could be an appropriate approximation). Additionally, the logarithmic distance effect is partly confirmed by the second experiment and by the inspection of the residuals (results not presented here).

Detecting the distance and size effects.

The present analysis is not relevant in contrasting the ANS and DSS models, but in the second and third experiments the existence of the numerical distance and size effects was tested, and the same analysis was run in the present experiment, to be able to use these results as a point of reference. The slopes of the specific effects were tested (1) with multiple linear regressions, and (2) with simple linear regressions.

Methods for multiple linear regression.

Average error rates and median reaction times of the correct responses were calculated for each number pair for each participant. Error rates and reaction times were fitted with two regressors for all participants: (a) distance effect (the absolute difference of the two values), (b) size effect (the sum of the two values). See the values of the regressors for the whole stimulus space on Figure 5. (The end effect regressor is used only in the second and third experiments.) This analysis gives a more stable result compared to the more commonly applied simple linear regression analysis (see below). The weights of the regressors were calculated for each participant in both error rates and reaction times, and all regressors' values were tested against zero.

Distance effect

	1	2	3	4	5	6	7	8	9
1	0	1	2	3	4	5	6	7	8
2	1	0	1	2	3	4	5	6	7
3	2	1	0	1	2	3	4	5	6
4	3	2	1	0	1	2	3	4	5
5	4	3	2	1	0	1	2	3	4
6	5	4	3	2	1	0	1	2	3
7	6	5	4	3	2	1	0	1	2
8	7	6	5	4	3	2	1	0	1
9	8	7	6	5	4	3	2	1	0

Size effect

	1	2	3	4	5	6	7	8	9
1	2	3	4	5	6	7	8	9	10
2	3	4	5	6	7	8	9	10	11
3	4	5	6	7	8	9	10	11	12
4	5	6	7	8	9	10	11	12	13
5	6	7	8	9	10	11	12	13	14
6	7	8	9	10	11	12	13	14	15
7	8	9	10	11	12	13	14	15	16
8	9	10	11	12	13	14	15	16	17
9	10	11	12	13	14	15	16	17	18

End effect

	1	2	3	4	5	6	7	8	9
1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1
2	0.5	0	0	0	0	0	0	0.5	1
3	0.5	0	0	0	0	0	0	0.5	1
4	0.5	0	0	0	0	0	0	0.5	1
5	0.5	0	0	0	0	0	0	0.5	1
6	0.5	0	0	0	0	0	0	0.5	1
7	0.5	0	0	0	0	0	0	0.5	1
8	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1
9	1	1	1	1	1	1	1	1	1

Figure 5. Values of the three regressors applied in the multiple linear regression in the whole stimulus space.

Methods for simple linear regression.

To test our data with a more commonly applied simple linear regression, all multiple linear regression analyses were retested. For the distance effect the trials were grouped according to distance (absolute difference between the two numbers) for all participants. For the size effect the trials were grouped according to the sum of the two numbers, excluding trials with distance larger than 3. The latter was necessary, because otherwise the specific shape of the stimulus space and the distance effect might cause an artifact size effect: Cells from the middle part of the size range include more large-distance cells than cells from the end part of the size range do. Linear slope was fitted both on the error rates and on the reaction times for both the distance and size effects for all participants, then the slopes were tested against zero. Because the simple linear regression analysis gave the very same pattern as the multiple linear regression for all experiments of the present work, the results of this analysis are not presented here.

Results and discussion.

Fitting the functions of the ANS and the DSS models to the reaction times.

For the reaction time analysis median reaction time of the correct responses for each number pair and for each participant was calculated. The mean of the participants data for all number pairs (Figure 6) were fit linearly with the least square method. Four models were fit to the group mean: The Moyer and Landauer version of the ANS function, the Crossman version of the ANS function, the linear DSS function, and the logarithm DSS function (see Methods for their descriptions).

For the Moyer and Landauer version the data showed a quite appropriate fit, with $R^2 = 0.884$, $AIC = 613.8$, while the Crossman version of the ANS function fit was somewhat worse, although similar, with $R^2 = 0.769$ and $AIC = 663.5$. Regarding the DSS models, the fit for the linear version was $R^2 = 0.808$, $AIC = 652.4$, and the fit for the logarithm version was $R^2 = 0.893$, and $AIC = 610.3$.

Overall, fitting the functions of the four versions of the two models resulted in similar *AIC*s within the same range, therefore no clear preference for any model can be pronounced. It seems that either the appropriate function is not precise enough to have a higher fit (which could be true for either the ANS or the DSS model), and/or with the current noise of the data the subtle differences between the models cannot be investigated. Thus, reaction time analysis with the current functions and the available signal-to-noise ratio could not be decisive in contrasting the ANS and DSS model.

Error rate

Number on the right (columns); Number on the left (rows)

	1	2	3	4	5	6	7	8	9
1		1%	0%	0%	1%	1%	0%	0%	0%
2	2%		2%	3%	0%	1%	3%	0%	1%
3	2%	1%		4%	2%	1%	0%	0%	1%
4	1%	1%	5%		7%	1%	4%	2%	1%
5	1%	1%	1%	4%		7%	7%	2%	2%
6	1%	1%	0%	3%	2%		9%	6%	2%
7	0%	2%	2%	4%	7%	12%		11%	6%
8	1%	0%	0%	0%	2%	6%	7%		6%
9	0%	0%	1%	0%	4%	3%	6%	11%	

Reaction time

Number on the right (columns); Number on the left (rows)

	1	2	3	4	5	6	7	8	9
1		494	487	481	462	449	462	444	448
2	502		556	538	503	489	491	474	469
3	480	513		591	537	517	527	487	483
4	474	523	560		576	515	555	498	505
5	459	487	528	560		577	589	524	521
6	457	467	497	513	588		633	560	554
7	457	493	509	530	576	624		605	559
8	447	473	472	488	518	600	580		622
9	457	456	483	482	514	561	541	604	

Drift rate

Number on the right (columns); Number on the left (rows)

	1	2	3	4	5	6	7	8	9
1		0.33	0.29	0.34	0.35	0.38	0.37	0.40	0.41
2	0.32		0.27	0.29	0.34	0.34	0.36	0.37	0.37
3	0.29	0.32		0.27	0.29	0.30	0.34	0.34	0.36
4	0.32	0.30	0.27		0.22	0.30	0.31	0.29	0.31
5	0.36	0.32	0.31	0.28		0.24	0.26	0.31	0.32
6	0.36	0.36	0.32	0.25	0.24		0.21	0.24	0.27
7	0.32	0.31	0.31	0.27	0.23	0.20		0.21	0.26
8	0.39	0.34	0.36	0.34	0.29	0.23	0.23		0.24
9	0-36	0.37	0.35	0.31	0.29	0.25	0.25	0.22	

Figure 6. Error rates (left), response times in ms (middle) and drift rates (right) in the Indo-Arabic digits number comparison for the whole stimulus space. Lighter shade denotes fast and error-free responses, darker shade denotes slow and erroneous responses. Results show distance and size effects.

Fitting the functions of the models to the error rates.

For the error rate analysis, the mean error rate for each number pair and for each participant was calculated, then the average of the participants was computed (Figure 6). To test the ANS model, first, we looked for the Weber ratio that gives the same mean error rate for the stimulus space used here (all possible number pairs for numbers between 1 and 9, ties excluded) as it was measured in our data (2.5%). The found 0.11 Weber ratio was used to generate the predictions of the ANS model for all cells of the stimulus space (see Methods for the function), and the model was linearly fit to the error rate data with the least square method. The goodness of fit was $R^2 = 0.625$, *AIC* = -371. In testing the DSS model, the goodness of fit for the linear version was $R^2 = 0.505$, *AIC* = -341, and the logarithmic DSS model gave a goodness of fit of $R^2 = 0.667$, *AIC* = -377.

Like in the case of the reaction time, the goodness of fit of the ANS and the DSS models are indistinguishable in the error rates data. This again shows that with the signal-to-noise ratio of the present data, the two models are indistinguishable, or the DSS model is not precise enough to show a higher fit.

Fitting the functions of the models to the drift rates.

To recover the drift rates for all number pairs in the two notations, the EZ diffusion model was applied, which can be used when the number of trials per cells is relatively small (Wagenmakers et al., 2007). For edge correction we used the half trial solution (see the exact details about edge correction in Wagenmakers et al., 2007). The scaling within-trials variability of drift rate was set to 0.1 in line with the tradition of the diffusion analysis literature. Drift rates for each number pair and participant were calculated in both notations. The mean drift rates of the participants for the full stimulus space are displayed in Figure 6.

According to the goodness of fit of the models, the ANS model is worse (*AIC* = -140.1) than the DSS model (*AIC* = -332.4 and *AIC* = -348.1 for the linear and logarithmic DSS model versions, respectively). (Because in a linear fit with zero intercept, the R^2 is much higher than in a linear fit with non-zero intercept (as a consequence of some of the mathematical properties of R^2), and because the ANS model uses 0 intercept, but the DSS model does not, the R^2 values are not reported here.)

Looking at the drift rates of the comparison task (Figure 6) might reveal why the ANS model is worse than the DSS model: While the ANS model predicts that the drift rate tends to zero as the stimuli become indistinguishable (e.g., 8 vs. 9), the recovered drift rates are in fact much larger, tending to the 0.2 values. This problem is analogous to a conceptual problem: How is it possible that an imprecise representation solves a precise comparison task? In other words, if the Weber fraction of the ANS is around 0.11, how is it possible that small ratio number pairs, e.g., 8 vs. 9, can still be differentiated with relatively high precision.

Thus, in the diffusion model analysis the DSS model seems to offer a better prediction than the ANS model, however, it is important to note that (a) the EZ diffusion model analysis and more generally any diffusion models have some constrains (Wagenmakers et al., 2007), and consequently, it is possible that in this case the recovered parameters are not entirely reliable, and (b) task difficulty can be defined in different ways (Dehaene, 2007; Palmer et al., 2005), and it might be debated which definition is appropriate. Thus, while the present diffusion model analysis reveals the advantage of the DSS model over the ANS model, the uncertainties of the methods might question how reliable these results are. (The methods and the models are investigated in more details in Krajcsi et al., 2018).

Presence of the distance and the size effects.

According to the multiple linear regression analysis, both the distance and the size effects were present both in the error rates and in the reaction times, 95% CI for the slope was [-1.16%, -0.65%], $t(17) = -7.42$, $p < 0.001$ for the distance effect in error rates, and CI of [-23.6 ms, -15.5 ms], $t(17) = -10.1$, $p < 0.001$ in reaction times, CI with [0.3%, 0.59%], $t(17) = 6.57$, $p < 0.001$ for the size effect in error rates, and CI with [4.8 ms, 9.1 ms], $t(17) = 6.78$, $p < 0.001$ in reaction times.

Summary.

First, we found that reaction time and error rate patterns in Indo-Arabic number comparison (Figure 6) could not be decisive in contrasting the ANS and the DSS models. Even if the two models correlate, the correlation is not perfect, and there was a chance that the present test could have decided. Still, with the present models and/or signal-to-noise ratio, the test was not decisive. On the positive side, this means that the DSS model is a viable alternative to the ANS model, because the goodness of fit of the DSS model is in the same range as the goodness of fit of the ANS model. Second, we found that in a diffusion model analysis the drift rate pattern is more in line with the DSS model than with the ANS model, although the uncertainties about the method may question the reliability of these results. Overall, while the performance in the Indo-Arabic comparison task suggests that the DSS model is a viable model, this paradigm could not decide firmly which model is preferred. Thus, in the next experiment a new approach is utilized in which we investigate the role of the frequency in the size effect.

Experiment 2 – Role of the Frequency in the Size Effect

In a different approach, we tested whether the distance and the size effects are strongly related as suggested by the ratio-based ANS model, or whether the two effects can dissociate. In the present experiment we investigated whether size effects can dissociate from distance effect if the frequency of the symbols is manipulated. (See another type of test for the dissociation of the two effects in Krajcsi, 2016) To manipulate the frequency of the symbols, it might be more appropriate to use new symbols, instead of the well-known Indo-Arabic symbols, because the frequency of the already known symbols might be well established and learned.

Thus, to investigate the role of the frequency in the size effect, participants learned new number symbols in a simple number comparison task, and the frequency of the symbols was manipulated in the experiment. According to the DSS model, the size effect could be changed as a function of the symbol frequencies (Figure 3), if the reaction time depends on the frequency of the symbol, and not the frequency of the concept. For example, if the distribution of the frequencies is uniform, then according to the DSS model, the size effect should vanish. In contrast, according to the ANS model, even with uniform frequency distribution the size effect should be visible, because the size effect is rooted in the ratio of the two values, independent of the frequency (Figure 7). It is important to stress that although according to the ANS model it might be possible that the frequency of the symbols have an effect on the performance, the effect should be relatively weak: Although in the ANS model the role of the frequency is not discussed, it states that the largest part of the performance variance should be explained by the ratio (Dehaene, 2007; Moyer & Landauer, 1967), which means that any other factors could have only a minor effect on the performance.

Methods.

Participants learned new symbols (Figure 8) for the numbers between 1 and 9 to compare, while the frequency of all symbols was manipulated in two conditions.

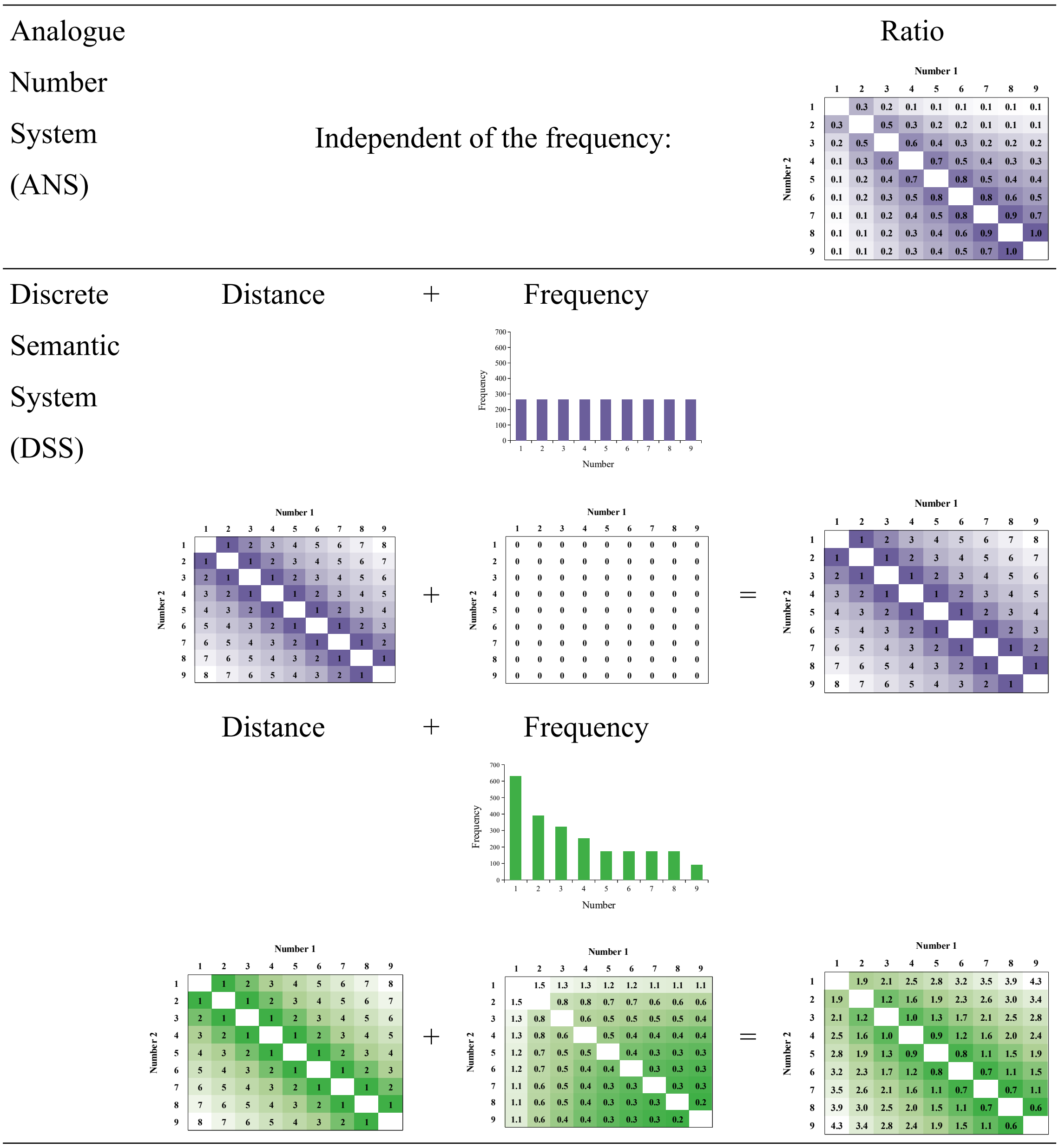


Figure 7. Prediction of the two models for the symbol frequency manipulation in Experiment 2. Bar charts show the frequency of the stimuli used in the uniform distribution condition and in the Indo-Arabic-like distribution condition. (In the Indo-Arabic-like distribution the resulting performance is computed as 0.4 × Distance + Frequency.)

It is possible that the new symbols are not connected to the numerical values they represent, and they may be processed only as a non-numerical ordered series. This could cause a problem, because the ANS could not process this non-numerical order[8].

8 Note, however, that several works suggest that order processing and quantity processing rely on the same mechanisms (Leth-Steensen & Marley, 2000; Marshuetz, Reuter-Lorenz, Smith, Jonides, & Noll, 2006; Verguts & Van Opstal, 2014), thus, ANS should be activated even when the new symbols are non-numerical orders.

To ensure that the new symbols were connected to the numerical values they represent, at the end of the experiment we used a priming task to measure the priming distance effect (PDE) between the newly learned symbols and familiar Indo-Arabic digits (Figure 8). In a PDE the reaction time to the target is faster when the numerical distance between the prime and the target is smaller, reflecting a semantic relation between the prime and the target (Koechlin et al., 1999; Reynvoet & Brysbaert, 1999; Reynvoet et al., 2009).

Participants.

Eighteen university students participated in the uniform frequency distribution condition. After excluding 2 of them because the error rate did not fall below 5% even after the 5th block, and excluding 2 further participants showing higher than 5% error rates in the main comparison task, the data of 14 participants was included (11 females, mean age 20.6 years, standard deviation 2.1 years). Fifteen university students participated in the Indo-Arabic-like frequency distribution condition. After excluding two participants because their error rates were higher than 5% either in the main comparison or in the priming comparison task, the data of 13 participants was analyzed (13 females, mean age 24.3 years, standard deviation 6.9 years).

Stimuli and procedure.

The participants first learned new symbols for the numbers between 1 and 9. Then, in a comparison task they decided which number is larger in a simultaneously presented symbol pair. Finally, in a priming comparison task the participants decided whether one-digit numbers are smaller or larger than 5 (Figure 8).

New symbols were introduced to represent values between 1 and 9. The new symbols were chosen from writing systems that were mostly unknown to the participants, and the characters had similar vertical and horizontal size. The symbols were randomly assigned to values for all participants, i.e., the same symbol could mean a different value for different participants, from the following characters: Ꮳ, Ꮘ, Ꭷ, Ꮷ, Ꮽ, Ł, Ɵ, Đ, И, Я, Ꮞ, Ħ, Œ, ƌ, ƛ, ʕ, Ʊ, Ʋ.

To ensure that the participants have learned the symbols, in the symbol learning phase, the symbols were practiced until a threshold hit rate was reached. In a trial of the new symbol learning phase a new symbol and an Indo-Arabic digit were shown simultaneously, and the participant decided whether the two symbols denote the same

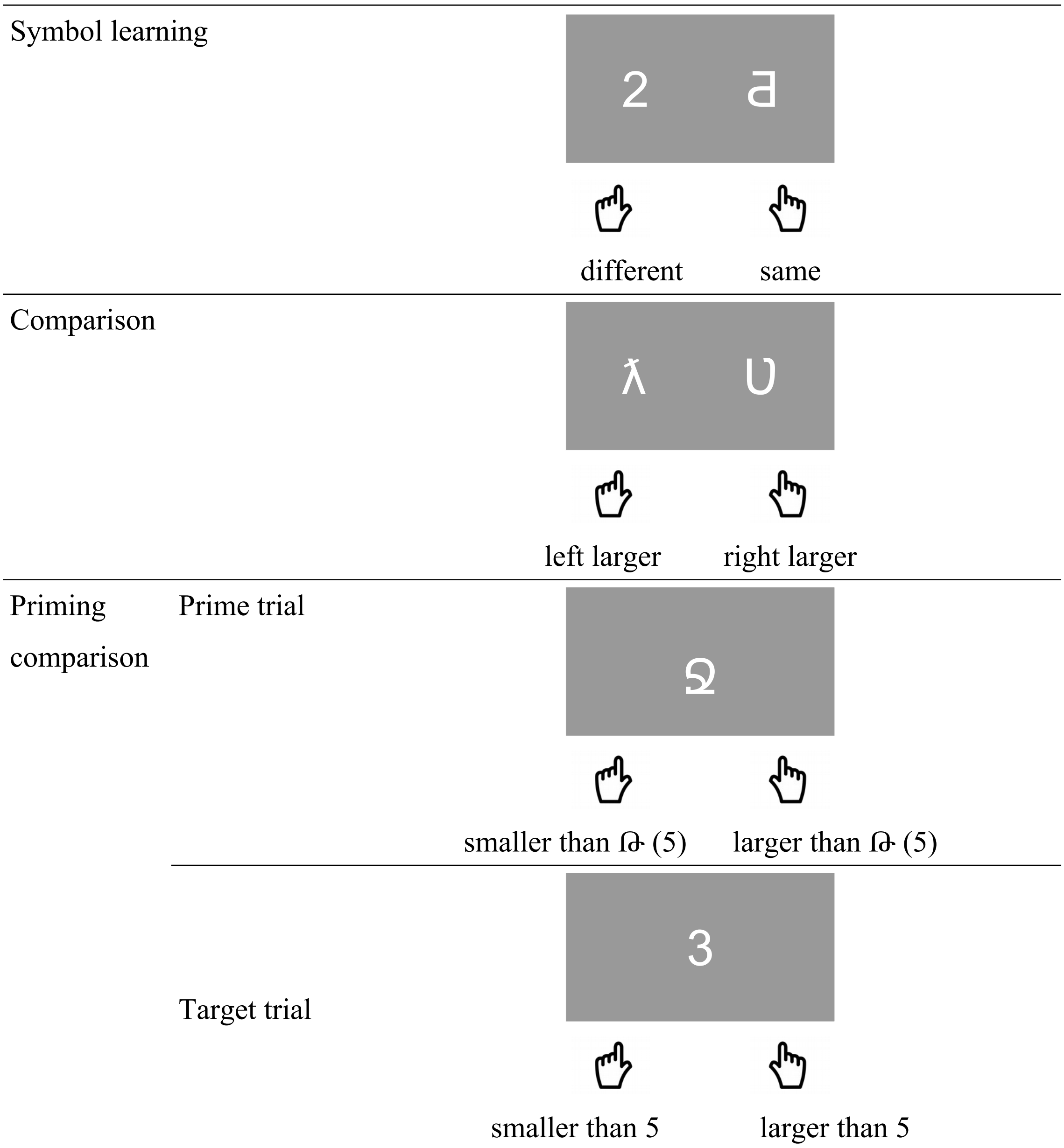


Figure 8. Tasks in the new symbol experiments.

value. The stimuli were visible until response. After the response, auditory feedback was given. In a block all symbols were presented 10 times (90 trials in a block) in a randomized order. In half of the trials the symbols denoted the same values. The symbol learning phase ended if the error rate in a finished block was smaller than 5%, or the participant could not reach that level in five blocks.

In the main comparison task, the same procedure was used as in the first experiment, but here the numbers were denoted with the new symbols. In the uniform frequency distribution condition the number of the presentation of a digit were the same as in the first experiment (all possible number pairs were shown 10 times). In the Indo-Arabic-like frequency distribution condition the frequencies of the specific values followed the frequencies of the numbers in everyday life (Dehaene & Mehler, 1992), specified with the following formula: $\text{frequency}_{\text{value}} = \text{value}^{-1} \times 10$. This formula generated the following frequencies (value:frequency): 1:10, 2:5, 3:4, 4:3, 5:2, 6:2, 7:2, 8:2, 9:1 (Figure 7). The 2-permutations of these numbers excluding ties were presented, resulting in 794 trials.

In the priming comparison task in odd (prime) trials a new symbol was visible, and the participant decided whether it was smaller or larger than 5. Two hundred ms after the response in an even (target) trial an Indo-Arabic digit was shown, and the participant decided whether it was smaller or larger than 5. Two thousand ms after the response a new odd (prime) trial was shown. The stimuli were visible until response. The instruction included the value of 5 in both notations: For even trials Indo-Arabic notation (“5”), for odd trials the new notation (e.g., “ꝭ”) was used. All possible new symbols were presented with all possible Indo-Arabic digits three times, resulting in 192 trials.

Results and discussion.

To summarize the main results, in the uniform distribution comparison task the distance effect was present, but the size effect was not (Figure 9A). This result is in line with the DSS model, but not with the ANS model. On the other hand, in the Indo-Arabic-like, biased frequency comparison task both the distance and the size effects were visible (Figure 9B) in a similar pattern as observable in Indo-Arabic number comparison (Figure 6), suggesting that it is the frequency manipulation that is responsible for the size effect.

Distance and size effects in the uniform frequency distribution.

The same analysis methods were applied as in the first experiment with two exceptions. Descriptive data clearly shows an end effect (Leth-Steensen & Marley, 2000). Thus, an end effect regressor was also included in the multiple linear regressions (Figure 5) with a value of 1 if any of the presented numbers were 9, 0.5 if any of the numbers were 8 or 1, and 0 otherwise. These values were specified with first calculating the average reaction time for all presented numbers, then the distance effect (distance from 5) of the middle number range (i.e., without end effect) was extrapolated, and finally, the deviation from this extrapolation at the end of the number range was estimated.

A Equal frequencies condition

Number on the right

Number on the left	1	2	3	4	5	6	7	8	9
1		4%	4%	2%	0%	0%	0%	0%	0%
2	6%		5%	1%	0%	0%	0%	0%	0%
3	1%	4%		2%	0%	1%	1%	0%	0%
4	2%	1%	1%		5%	2%	2%	0%	1%
5	1%	0%	2%	4%		2%	1%	0%	0%
6	0%	0%	1%	2%	4%		7%	1%	3%
7	0%	0%	0%	2%	2%	6%		7%	2%
8	0%	0%	0%	1%	0%	1%	2%		4%
9	0%	0%	0%	0%	0%	0%	1%	4%	

Number on the right

Number on the left	1	2	3	4	5	6	7	8	9
1		1274	1196	1026	989	926	1088	844	810
2	1283		1309	1275	1163	1153	1085	1000	885
3	1270	1432		1591	1292	1261	1141	1022	888
4	1176	1354	1556		1381	1328	1202	1024	887
5	1066	1225	1326	1502		1387	1211	1124	908
6	998	1134	1278	1343	1358		1435	1240	956
7	1075	1153	1230	1237	1247	1463		1229	960
8	945	1003	1009	1147	1040	1288	1179		1008
9	782	825	821	823	780	885	846	911	

B Biased frequencies condition

Number on the right

Number on the left	1	2	3	4	5	6	7	8	9
1		2%	1%	1%	0%	0%	0%	0%	0%
2	4%		5%	1%	1%	0%	2%	2%	0%
3	0%	5%		5%	2%	2%	0%	1%	2%
4	0%	2%	2%		1%	1%	3%	0%	0%
5	0%	2%	4%	4%		13%	8%	4%	0%
6	0%	0%	0%	8%	19%		8%	0%	0%
7	0%	1%	0%	1%	4%	6%		13%	0%
8	0%	0%	1%	0%	4%	10%	13%		19%
9	0%	0%	2%	0%	8%	0%	4%	0%	

Number on the right

Number on the left	1	2	3	4	5	6	7	8	9
1		856	815	822	811	817	800	784	737
2	926		1206	1127	1168	1174	1117	937	885
3	880	1234		1451	1287	1322	1190	1094	924
4	923	1191	1411		1508	1461	1270	1222	896
5	882	1171	1359	1484		1756	1362	1113	994
6	871	1241	1299	1202	1845		1569	1184	1037
7	818	1054	1237	1362	1370	1554		1342	1037
8	843	937	989	1257	1169	1131	1272		1002
9	770	819	860	847	935	1274	1150	1098	

*Figure 9. Error rates (left) and response times in ms (right) in the new symbol number comparison for the whole stimulus space. Lighter shade denotes fast and error-free responses, darker shade denotes slow and erroneous responses. **(A)** Equal frequencies condition, showing distance and end effects. **(B)** Biased frequencies condition, showing distance, size and end effects.*

In the multiple linear regression the slope of the distance effect deviated from zero, 95% CI was [-1.04%, -0.48%], $t(13) = -5.84$, $p < 0.001$ for error rates, and CI was [-73.6 ms, -26.1 ms], $t(13) = -4.53$, $p = 0.001$ for reaction time. On the other hand, the slope of the size effect did not differ from zero, CI with [-0.15%, 0.06%], $t(13) = -0.933$, $p = 0.368$ for error rates, and CI with [-26.6 ms, 13.9 ms], $t(13) = -0.679$, $p = 0.509$ for reaction time. The end effect was present for the reaction time, CI of [-430.6 ms, -147.6 ms], $t(13) = -4.41$, $p = 0.001$, and more unstably for the error rates, CI with [-0.23%, 2%], $t(13) = 1.71$, $p = 0.111$.

These results also demonstrated an end effect (the most extreme values in the set are easier to respond than other values) (Leth-Steensen & Marley, 2000), however, while the end effect can be in line with the DSS model (Leth-Steensen & Marley, 2000), it is also possible that the effect is irrelevant in the description of the representation processing the numerical values (Balakrishnan & Ashby, 1991; Piazza, Giacomini, Le Bihan, & Dehaene, 2003), consequently, the presence of this effect is not decisive in the present question.

Distance and size effects in the Indo-Arabic-like frequency distribution.

The slope of the distance effect differed from zero in both the error rates, CI with [-1.56%, -0.5%], $t(12) = -4.25$, $p = 0.001$, and in reaction times, [-55.7 ms, -28.9 ms], $t(12) = -6.87$, $p < 0.001$. The non-zero slope of the size effect was also observable, [0.20%, 0.68%], $t(12) = 3.99$, $p = 0.002$ for the error rate, and CI with [28.4 ms, 50.2 ms], $t(12) = 7.85$, $p < 0.001$ for the reaction time. Additionally, the end effect was observable in the reaction times, CI with [-622.5 ms, -294.9 ms], $t(12) = -6.1$, $p < 0.001$, but not in the error rates, CI with [-2.76%, 0.7%], $t(12) = -1.3$, $p = 0.217$.

We tested directly whether the size effects of the two frequency conditions differed. The size effect slopes between the uniform frequency distribution and the Indo-Arabic-like frequency distribution conditions differed significantly in both the error rates, $U = 13$, $p < 0.001$, and in the reaction times, $U = 15$, $p < 0.001$.

Priming distance effect.

In this analysis the error rates and median reaction times of the correct responses of the target Indo-Arabic numbers were analyzed as a function of the prime (new symbols) – target (Indo-Arabic digit) distance (Figure 10). Only the trials in which the response was the same for the prime and distance (i.e., both numbers were smaller than 5, or both numbers were larger than 5) were analyzed (Koechlin et al., 1999; Reynvoet & Brysbaert, 1999; Reynvoet et al., 2009). Linear slope was calculated for the PDE.

In the uniform frequency distribution the data of one participant was not recorded due to technical problems. Because in the symbol learning task participants practiced the new symbol – Indo-Arabic pairs, the zero distance pairs could have this extra practice gain, and not depend purely on the semantic priming effect. Thus, the 0 distance pairs were not included in the analysis. While the descriptive data showed increasing priming effect with smaller distance (Figure 10), the effect was not significant: In the uniform frequency condition CI is [-1.62%, 2.69%], $t(12) = 0.54$, $p = 0.599$ for the error rate, and CI is [-1.4 ms, 39.7 ms], $t(12) = 2.03$, $p = 0.065$ for the reaction time, and in the Indo-Arabic frequency condition CI is [-0.13%, 1.63%], $t(12) =$

1.85, $p = 0.089$ for the error rates, and CI is [-8.9 ms, 27.2 ms], $t(12) = 1.11$, $p = 0.290$ for the reaction time. The lack of significance could mean the lack of PDE, or it could reflect the lack of statistical power, or both. Looking at the gradual increase of error rate and reaction time as the function of priming distance (Figure 10) and the biased CIs, it seems more probable that the PDE could be statistically significant with larger statistical power. To extend the reasoning that the lack of the significance might be the result of insufficient statistical power, we also analyzed three unpublished similar experiments conducted in our laboratory, where in the same design new symbols were learned with the same stimuli and procedure as in the present works (in the third unpublished experiment the learning and the comparison were repeated for 5 days). In those experiments the PDE was measured with similar sample sizes as in the experiments presented here. We found that in all cases the confidence interval was biased to the direction the PDE predicts, although mostly it was only close to be significant. In the first two unpublished experiments 95% CI is [-6.2 ms, 17.7 ms], $N = 12$, $p = 0.312$, and [29.1 ms, 75.1 ms], $N = 10$, $p < 0.001$. In the third unpublished experiment the PDE was measured for 5 consecutive days which is especially informative about the consistency and fluctuation of the PDE in a relatively small sample: 95% CI [10.25 ms, 34.54 ms], $N = 13$, $p = 0.002$, [-2.09 ms, 28.21 ms], $p = 0.085$, [-1.05 ms, 14.39 ms], $p = 0.084$, [-9.34 ms, 8.13 ms], $p = 0.882$, [-0.49 ms, 13.24 ms], $p = 0.066$, for the five days, respectively. A meta-analysis on the five available experiments (second and third experiments of the present paper and three unpublished experiments; only day 1 was used from the last unpublished experiment; meta-analysis of means in original units with random effect) revealed 95% CI [6.7 ms, 34.3 ms], $p = 0.004$ (Cumming, 2012). The analysis also confirms that the effect size would require much larger sample to have a significant result reliably in a single experiment: The estimated effect size could be as small as $d = 0.3$ (with around 25 ms standard deviation), which would require a magnitude of 100 participants to reach 95% statistical power (Faul et al., 2007). Taken together, based on (a) the expected gradual pattern of the PDE (Figure 8), and (b) the consistently biased CIs across experiments, (c) confirmed with the meta-analysis, it is most reasonable to conclude that the PDE is present, even if our usual sample size around 15 does not guarantee the preferred 95% statistical power for a single experiment.

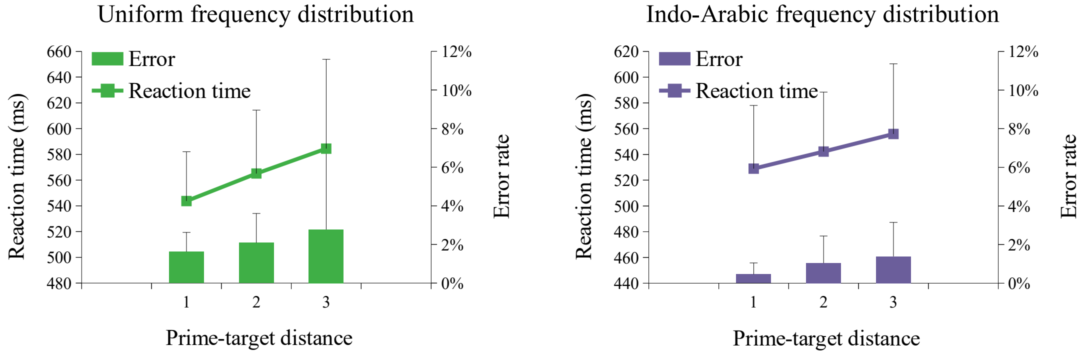


Figure 10. *Prime distance effect (PDE) measured in error rates (bars) and reaction time (lines), in equal frequency condition (left) and in biased frequency condition (right) in Experiment 2. Error bars represent 95% confidence interval.*

Effect of the frequency.

To further demonstrate the effect of the frequency (because it cannot be observed readily on Figure 9), the mean reaction time was calculated for all cells that include a specific value in both conditions (right of Figure 11). The reaction time changes in line with the frequencies of the values: The more frequent a number is in one condition compared to the other condition (left of Figure 11), the faster it is to process (right of Figure 11). In other words, the differences of the two conditions for the values in the reaction time data are inversely proportional to the differences of the two conditions for the values in the frequency. Note that the reaction time data do not include purely the frequency effect, because (a) middle values are gradually slower to process because of the interaction of the distance effect and the shape of the stimulus space, and (b) end values are faster to process because of the end effect.

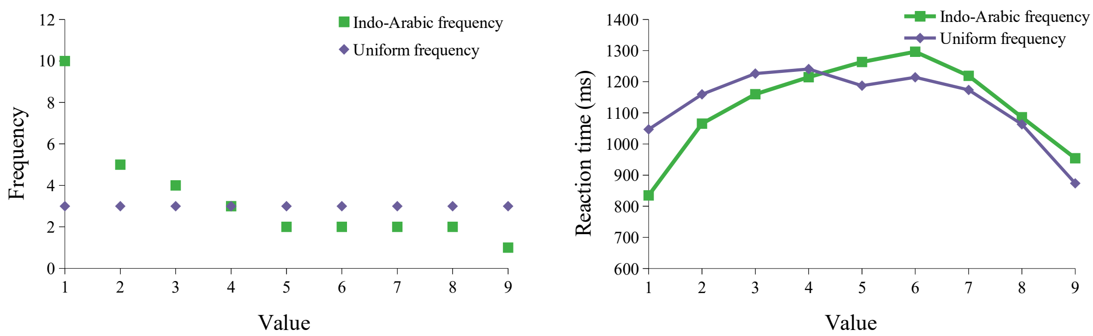


Figure 11. *Frequencies of the specific values (left) and response latencies for those values (right) in Experiment 2.*

Summary.

In the second experiment the numerical distance and size effects dissociated. More specifically, the numerical size effect was missing when the frequency distribution was uniform, and the size effect was present with the biased frequencies of symbols, suggesting that size effect was guided by the frequencies of the symbols. These results cannot be explained by the ANS model, whereas they can be in line with the DSS model. We highlight again that according to the ANS model although the frequency might slightly modulate the performance, it cannot change a large proportion of the variance in the performance. However, the present result reveal that largest part of the variance of the size effect is directed by the frequency, while the ratio has no observable effect (as revealed by the statistical lack of the size effect), contradicting the ANS model prediction.

Results also show that the new numbers semantically primed the Indo-Arabic digits as revealed by the PDE, demonstrating that the new symbols were connected to the values they represent. Thus, the lack of the size effect in the second experiment cannot be the result of potentially non-numeric new symbols which could not be processed by the ANS.

Experiment 3 – Role of the Semantic Congruency Effect in the Size Effect

As another potential confound, it is possible that in the second experiment there was a size effect in the uniform distribution condition, however, a semantic congruency effect (SCE) extinguished it. According to the SCE, large numbers are responded to faster than small numbers when the task is to choose the larger number, resulting in a reversed size-like effect, and small numbers are faster to decide on when the smaller number should be chosen, resulting in a regular size-like effect (Leth-Steensen & Marley, 2000). If the SCE was present in the second experiment, this anti-size effect could have extinguished a potentially existing size effect. To test this possibility, the uniform frequency condition of the second experiment was rerun, but this time the participants had to choose the smaller number. If the SCE was present in the second experiment as a reversed size-like effect, then it should be observed in the present experiment as a regular size-like effect, increasing the size effect. However, the size effect was not present in this control experiment, demonstrating that the SCE did not mask a potentially existing size effect.

Methods.

The methods of the second experiment was applied, however, participants had to choose the smaller number, not the larger, in the comparison task. The priming comparison task was not run.

Eighteen university students participated in the study. Two participants were excluded, because their error rates were higher than 5% after the 5th learning block, and two participants were excluded because they had higher than 5% error rate in the comparison task. The data of 14 participants were analyzed, 10 females, with mean age of 25.4 years, and standard deviation of 6.9.

Results.

Distance and size effects.

In the multiple linear regression analysis the distance effect was present in both the error rate and the reaction time, CI [-1.54%, -0.48%], $t(13) = -4.13$, $p = 0.001$, and CI [-77.0 ms, -39.6 ms], $t(13) = -6.73$, $p < 0.001$, respectively. More critically, the size effect was not observable neither in the error rate nor in the reaction time, CI [-0.18%, 0.15%], $t(13) = -0.184$, $p = 0.857$, and CI [-25.0 ms, 26.2 ms], $t(13) = 0.0482$, $p = 0.962$, respectively. Comparing the slopes of the uniform frequency condition of the second and the present experiments, the slopes of the size effects did not differ significantly, neither in error rate nor in reaction time, $t(26) = 0.33$, $p = 0.744$, and $U = 91.5$, $p = 0.783$, respectively. Thus, choosing the smaller number did not change the size effect, consequently, the SCE did not influence essentially the size effect in the second experiment.

General Discussion

We introduced the DSS model as a comprehensive alternative account to the ANS model to explain symbolic number processing. First, we have shown that the DSS model can explain many symbolic numerical effects, and we demonstrated that the DSS model could give a similar quantitative prediction for symbolic number comparison performance as the ANS model. Second, we tried to contrast the two models in Indo-Arabic comparison task. However, because of the relatively high noise and the uncertainties of the diffusion analysis method, it was not possible to find a straightforward preference for any models. On the other hand this result could show that the DSS model prediction empirically fits the Indo-Arabic number comparison as good as the ANS model prediction. Finally, the second and third experiments revealed that in new symbol comparison tasks the numerical size effect is the consequence of the frequency manipulations of the symbols, as proposed by the DSS model, and not the consequence of the ratios of the values, as predicted by the ANS model. These data also show that the numerical distance and size effects are not straightforward signs of the ANS, because an alternative mechanism could produce them as well.

While the second and the third experiments utilized new symbols, it is possible to extend our conclusion about other symbolic number comparisons, for example, the Indo-Arabic number comparison. Because all known numerical effects that were observable in the new symbol comparison show the very same pattern as in Indo-Arabic comparison (i.e., distance effect, size effect and PDE), it is parsimonious to suppose that the same mechanisms work behind new symbol comparison and Indo-Arabic comparison, and our findings can also be generalized to the Indo-Arabic and other symbolic number processing, unless additional data show the opposite.

We argue that the ANS model is not in line with our results. While one can try to modify the ANS model to align with the present result, ratio-based performance is a defining feature of the ANS, and changing that feature leads not only to a modified ANS model, but to a completely new model. Additionally, adding frequency effect to the ANS model cannot modify it to explain the frequency-based size effect, because the ANS critically suggests that the performance should mainly be driven by the ratio, which ratio effect in fact was statistically invisible in the second and third experiments.

We argue that the ANS model is not in line with the present results, and the DSS can be an appropriate alternative. However, one might question how strongly our results support a DSS model. Obviously, one can only tell if a model is in line with the empirical results, and whether the model is coherent. We argue that the DSS is in line with the present and previous results (e.g., it can explain the independent distance and size effects, why symbolic and non-symbolic comparisons are relatively independent, or how arbitrarily precise comparison can be made), and it is a coherent model. Additionally, based on current cognitive models, it is reasonable to suppose that abstract symbolic operations are processed by a system that is otherwise known to be used for other symbolic operations, such as the mental lexicon or a conceptual network. On the other hand, no one can exclude that an alternative, third model could account for these results, and not the ANS or the DSS models. Further research can tell whether the DSS framework is an appropriate explanation for the symbolic number processing or another alternative should be found. Furthermore, it is possible that it is not a single representation that is responsible for the discussed effects, but cooperation of several representations is required, and although the ANS cannot explain the distance effect in comparison task, still there could be other symbolic numerical phenomena that could be rooted in the ANS. Additional works can find out whether such a partial role can be attributed to the ANS in symbolic number processing.

The DSS model in its current form relies on models about mental lexicon or conceptual networks. These starting points could offer many properties of the models, while at the same time, many other details are seemingly missing, e.g., the exact quantitative description of the comparison performance. While these shortcomings might make the impression that the DSS model is less detailed than the ANS model, these differences are the consequence of changing the base of the explanations. While the ANS model is a low-level perceptual model in its nature, the DSS model is more like a linguistic or conceptual network model. Models describing higher level functions are usually less quantitative than models describing lower level functions, partly because of methodological reasons, and from this viewpoint it seems reasonable that a DSS model is less quantitative than an ANS model. However, from a different—and more relevant—viewpoint, the DSS model is as efficient as the ANS model, because seemingly all relevant symbolic numerical effects and phenomena can also be explained in the DSS model, and a few examples can already be found where the DSS model can give a better explanation than the ANS model.

The ANS model is a widely accepted and deeply grounded explanation for number processing. However, despite the huge amount of papers discussing and supporting the ANS view, they are relying on surprisingly few effects and findings that demonstrate an ANS activation. In fact, the few phenomena can also be explained in the alternative DSS model as well. Additionally and more importantly, an increasing number of findings are not in line with the ANS model. For example, symbolic and non-symbolic performance seems to be independent on many behavioral (Holloway & Ansari, 2009; Sasanguie et al., 2014; Schneider et al., 2017) and neural level (Bulthé et al., 2014; Bulthé et al., 2015; Damarla & Just, 2013; Lyons et al., 2015). In a correlational study it has been shown that distance and size effects dissociate in Indo-Arabic comparison task (Krajcsi, 2016). Some results show that the numerical representation is not analog: Functional activation in the brain while processing symbolic numbers seems to be discrete (Lyons et al., 2015), and symbolic numbers can also interfere with the discrete yes-no responses (Landy, Jones, & Hummel, 2008). The present finding showing the frequency dependence of the size effect also extends the list of results contradicting the ANS model. Future research can tell whether the ANS can be reformulated to account for these findings, or an alternative model, such as the DSS, can characterize symbolic number processing better.

Author Contributions

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

We thank Réka Eszter Börcsök, Dávid Czikora, József Fiser, Ildikó Király, and Krisztina Peres for their comments on an earlier version of the manuscript. The study was supported by the Institute of Psychology, Eötvös Loránd University.

Thesis Study 2

Symbolic Numerical Distance Effect Does Not Reflect the Difference between Numbers

Attila Krajcsi[1] and Petia Kojouharova[1,2]

[1]Department of Cognitive Psychology, Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary

[2]Doctoral School of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary

Article published online in *Frontiers in Psychology* on 26th November 2017

DOI: https://doi.org/10.3389/fpsyg.2017.02013

Both authors contributed equally to this work.

In a comparison task, the larger the distance between the two numbers to be compared, the better the performance—a phenomenon termed as the numerical distance effect. According to the dominant explanation, the distance effect is rooted in a noisy representation, and performance is proportional to the size of the overlap between the noisy representations of the two values. According to alternative explanations, the distance effect may be rooted in the association between the numbers and the small-large categories, and performance is better when the numbers show relatively high differences in their strength of association with the small-large properties. In everyday number use, the value of the numbers and the association between the numbers and the small-large categories strongly correlate; thus, the two explanations have the same predictions for the distance effect. To dissociate the two potential sources of the distance effect, in the present study, participants learned new artificial number digits only for the values between 1 and 3, and between 7 and 9, thus, leaving out the numbers between 4 and 6. It was found that the omitted number range (the distance between 3 and 7) was considered in the distance effect as 1, and not as 4, suggesting that the distance effect does not follow the values of the numbers predicted by the dominant explanation, but it follows the small-large property association predicted by the alternative explanations.

The Numerical Distance Effect and Its Explanations

In a symbolic number comparison task, performance is better (i.e., error rates are lower and reaction times are shorter) when the numerical distance is relatively large, e.g., comparing 1 vs. 9 is easier than comparing 5 vs. 6 (Moyer & Landauer, 1967). There are several explanations for this phenomenon termed the numerical distance effect.

According to the dominant model, numbers are stored on a continuous (analog) and noisy representation called the Analog Number System (ANS). The numbers are stored as noisy signals, and the closer the two numbers on the ANS, the larger the overlap of the two respective signal distributions is. As comparison performance is better when the overlap is relatively small, the large distance number pairs are easier to process because of the smaller overlap between the signals (Dehaene, 2007). More specifically, the ANS works according to Weber's law; therefore, the comparison performance depends on the ratio of the two numbers to be compared (Moyer & Landauer, 1967). In fact, the distance effect is the consequence of this ratio effect because larger distance also means higher ratio. The ratio effect is also thought to be the cause of the numerical size effect: Comparison performance is better for smaller numbers than for larger numbers because smaller number pairs have larger ratio than larger number pairs with the same distance (Moyer & Landauer, 1967). The ANS is thought to be the essential base of numerical understanding (Dehaene, 1992), and numerical distance effect is believed to be a diagnostic signal of the ANS activation while solving a numerical task.

However, there could be another explanation for the cause of the distance effect. Recently, it has been proposed that symbolic numerical effects, such as the distance and size effects, can be explained by a representation similar to the mental lexicon or conceptual networks, where nodes of the network represent the digits, and connections between them are formed according to their semantic and statistical relations (Krajcsi et al., 2016). In this model, termed the Discrete Semantic System (DSS) model, the numerical distance and size effects are rooted in two different mechanisms, even if the combination of these effects looks similar to the formerly supposed ratio effect. According to the model, the size effect might depend on the frequencies of the numbers: Smaller numbers are more frequent than larger numbers (Dehaene & Mehler, 1992); therefore, smaller numbers are easier to process, producing the numerical size effect. A similar frequency-based explanation of the size effect could be found in the model of Verguts et al. (2005). At the same time, numerical distance effect could be based on the relations of the numbers, for example, similar to the phenomenon in a picture naming task, where priming effect size depended on the semantic distance between the prime and target pictures (Vigliocco et al., 2002). There are several other alternative number processing models with partly overlapping suppositions and predictions as the DSS model (Leth-Steensen et al., 2011; Nuerk, Iversen, & Willmes, 2004; Pinhas & Tzelgov, 2012; Proctor & Cho, 2006; Verguts & Fias, 2004; Verguts et al., 2005; Verguts & Van Opstal, 2014). See the comparison of those models in Krajcsi et al. (2016) and in Krajcsi et al. (2018). Supporting the alternative DSS model, it has been found that the size effect followed the frequency of the digits in an artificial number notation comparison task (Krajcsi et al., 2016). In addition, it has been shown in a correlational study that in symbolic number comparison task, the distance and the size effects were independent (Krajcsi, 2016), reflecting two independent mechanisms generating the two effects. (See a similar prediction for independent distance and size effects in Verguts et al., 2005; Verguts & Van Opstal, 2014).

Because of the DSS model and the empirical findings demonstrating that the size effect is a frequency effect and that the distance and size effects are independent, it is essential to reconsider how the distance effect is generated. According to the DSS model, different explanations consistent with the supposed network architecture are feasible. First, it is possible that based on the values of the numbers, connections with different strengths between the numbers are formed—numbers with closer values have stronger connections—and stronger connections create interference in a comparison task, thereby resulting in a distance effect. This explanation is similar to the ANS model in a sense that value-based semantic relations are responsible for the distance effect. As an alternative explanation, it is also possible that based on previous experiences, numbers are associated with the "small" and the "large" properties, e.g., large digits, such as 8 or 9, are more strongly associated with "large," and small digits, such as 1 or 2, are more strongly associated with "small." These associations could influence the comparison decision, and the number pairs with larger distance might be easier to process because the associations of the two numbers with the small-large properties differ to a larger extent. A similar explanation has been proposed earlier in a connectionist model, which model predicted several numerical effects successfully, and one key component of this model was that the distance effect relies on the connection between the number layer and the "larger" nodes, where relatively large numbers are associated with the "larger" node more strongly than relatively small numbers (Verguts et al., 2005).

Therefore, the explanations of the numerical distance effect suppose two different sources for the effect: According to the ANS model and to the value-based DSS explanation, the effect is rooted in the *values or the distance of the numbers*, whereas in the association-based DSS explanation and in the connectionist model, the effect is rooted in the *strength of the associations between the number and the small-large properties*. The two explanations are not exclusive; it is possible that both information sources contribute to the distance effect.

The two critical properties of the two explanations, i.e., the values or distance of the numbers and the association between the numbers and the small-large properties, strongly correlate in the number symbols used in everyday numerical tasks. Therefore, in those cases, one cannot specify their role in the distance effect. However, in a new artificial number notation, the two factors (the distance of the values and the association) could be manipulated independently. This is only possible if the distance effect is notation specific. Otherwise, the new symbols would get the association strengths of the already known numbers, instead of forming new association strengths between the new symbols and the small-large properties. It is possible that the numerical effects are notation specific, as has been already demonstrated in the case of the numerical size effect: In an artificial number notation comparison task, the size effect followed the frequency of the digits, which also means that the size effect is notation specific (Krajcsi et al., 2016).

The Aim of the Study

The present study investigates whether in a new artificial number notation, where the values of the digits and the small-large associations do not necessarily correlate, the distance effect is influenced by the distance of the values or by the small-large associations, or both. One way to dissociate the two properties is to use a number sequence in which some of the values are omitted (Figure 12). If the distance effect is directed by the distance of the values, then the measured distance effect should be large around the gap (in this example, the effect should be measured as 4 units large), whereas if the distance effect is directed by the small-large associations, then the measured distance effect should be small around this gap, which is measured as an effect with a single unit distance, thereby supposing that the new digits were used in a comparison task with equal probability. If both mechanisms contribute to the distance effect, then the distance effect should be measured somewhere between the single unit and the many units (in this example 4 units) distance.

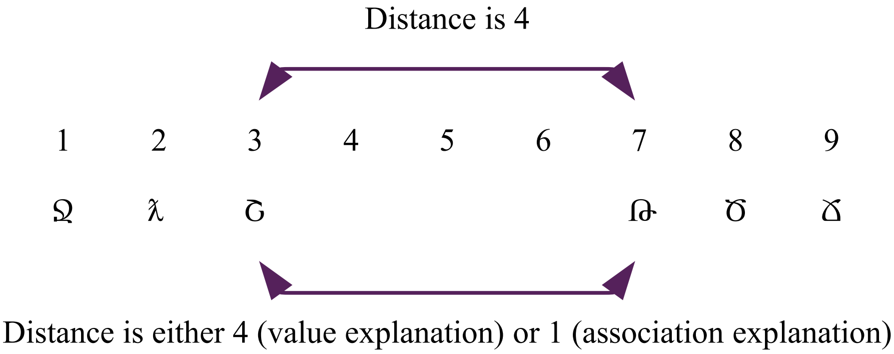


Figure 12. *An example of the symbols and their meanings in the present study. Arrows show the predicted distance effect size based on the predictions of the two explanations.*

Why does the association explanation predict a distance effect of 1 distance around the gap? In a comparison task, the association between a digit and the small-large properties may depend on how many times the digit were judged as smaller or larger. If the new digits are used with equal probability in the comparisons (and if the distance effect is notation specific), then the probability of being smaller or larger than the other number can be specified easily (see Table 2). In our example (Figure 12), the number 1 is always smaller. Hence, the association frequency is 100% with the small property and 0% with the large property. The number 2 is smaller when compared with 3, 7, 8, and 9, and larger when compared with 1. Therefore, the association frequency is 80% small and 20% large. Continuing the example, the association frequency is directly proportional to the order of the symbols and not to their value. If the distance effect depends on the order, then the distance between 3 and 7 (i.e., the two digits around the gap) is the same as any other neighboring digits (see the specific values in Table 2).

The two explanations predict different effect sizes for the distance effect not only for the two numbers next to the gap (e.g., for 3 vs. 7 on Figure 12 and Table 2) but also for any number pairs in which the two numbers are on the opposing side of the gap. The possible number pairs of the new symbols seen on Figure 12 and their hypothetical distance effect sizes according to the two explanations can be seen on Figure 13. Columns and rows denote the two numbers to be compared, and the cells show the

Table 2. The chance of being smaller or larger in a comparison task when the symbols are presented with equal probability.

Example symbols	Ω	ƛ	Ꮆ		Ꭷ	Ծ	Ճ
Meaning of the symbols	1	2	3		7	8	9
Chance of being smaller in a comparison	100%	80%	60%		40%	20%	0%
Chance of being larger in a comparison	0%	20%	40%		60%	80%	100%

distances of the value pairs (darker cells mean smaller distance). In the value explanation (left side), the predicted distance is the difference of the two numbers, whereas in the association explanation (right side), the predicted distance is computed based on the strength of the association with the small-large properties when the numbers are presented with equal probability, which is simply the order of those symbols in that series. The comparison performance should be proportional to the distance. Therefore, these figures show the performance pattern predictions according to the two explanations. The results will be displayed in a similar way as seen here because (a) displaying the full stimulus space is more informative than other indexes of distance effects, as any systematic deviation from the expected patterns could be observed, and (b) with the relatively large number of cells, any systematic pattern could be a convincing and critical information independent of the statistical hypotheses tests.

In the present test, it is critical that the new symbols should represent their intended values and not as a series that is independent of the intended number meanings; otherwise, the participants could consider the new symbols as numbers, e.g., from 1 to 6 because of their order in the new symbol series, which in turn could generate the performance predicted by the association explanation, even if the effect would be based on their values. One way to ensure that the new symbols are sufficiently associated to their intended values is to ensure that the priming distance effect works between the new and a well-known (for example, Indo-Arabic) notation. In numerical comparison tasks, the decision about the actual trial might be influenced by the stimulus of the previous trial, and the size of the influence is proportional to the numerical distance of the previous and actual stimuli, which is termed as the priming distance effect (PDE; Koechlin et al., 1999; Reynvoet & Brysbaert, 1999). The PDE is considered to be a sign of the relation between the symbols or the overlap of their

Value explanation

Number 2 (columns) × Number 1 (rows)

Number 1 \ Number 2	1	2	3	7	8	9
1		1	2	6	7	8
2	1		1	5	6	7
3	2	1		4	5	6
7	6	5	4		1	2
8	7	6	5	1		1
9	8	7	6	2	1	

Association explanation

Number 2 (columns) × Number 1 (rows)

Number 1 \ Number 2	1	2	3	4	5	6
1		1	2	3	4	5
2	1		1	2	3	4
3	2	1		1	2	3
4	3	2	1		1	2
5	4	3	2	1		1
6	5	4	3	2	1	

Figure 13. *The expected distance effect pattern for the stimulus space used in the present study based on the value explanation (left side) and based on the association explanation (right side). Specific values in the cells are the difference of the values (value model) or the difference of the order (association model) of the numbers to be compared on an arbitrary scale. Darker color indicates worse performance.*

representations (Van Opstal et al., 2008). Earlier experiments have shown that new artificial symbols can cause PDE in Indo-Arabic numbers (Krajcsi et al., 2016), suggesting that the new digits are not a series of symbols independent of their intended values, but they can be considered as a notation for the respective numbers. In the ANS framework, the PDE reflects the representational overlap between the numbers; thus, the PDE demonstrates that both notations appropriately activate the same representation – the ANS.

To summarize, the present study investigates whether the distance effect follows the distance of the values of the numbers (left of Figure 13) or the association of the small-large properties (right of Figure 13) or both, in the case of a newly learned notation (Figure 12), where some of the symbols are omitted. If both explanations are true, then we expect a pattern in-between the two figures, i.e., we should observe a break between 3 and 7 similar to the value explanation. However, the difference between the two sides of the gap should not be as large as in that explanation. All of these predictions only hold if the distance effect is notation specific; otherwise, the distance effect reflects the already well-known numbers, where the value and the association strongly correlates, and the pattern seen on the value model prediction can be expected. Consequently, only a pattern seen on the right in Figure 13 can decide about the models, because a pattern seen on the left can either mean a value-based distance effect or it can mean that the distance effect is notation independent.

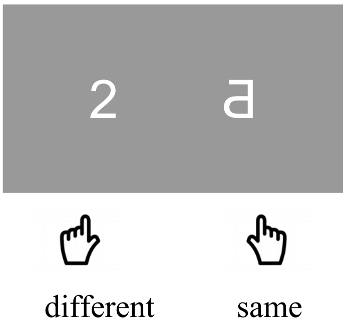
Methods

In the present experiment, participants learned new symbols (Figure 14), with the meaning of the numbers between 1 and 3, and between 7 and 9 (Figure 12). Then a number comparison task was performed with the new symbols (Figure 14).

Stimuli and procedure.

The new symbols were chosen from writing systems that were mostly unknown to the participants (e.g., ƌ, ƛ, Ʊ, Ջ). The characters had similar vertical and horizontal size, and similar visual complexity, and the height of the symbols were ~2 cm. (As mostly the apparent size does not influence the effects we investigate here, the visual angle was not controlled strictly.) Numbers were displayed in white on gray background. The symbols were randomly assigned to values for all participants, i.e., the same symbol could mean a different value for different participants.

Symbol learning



Comparison

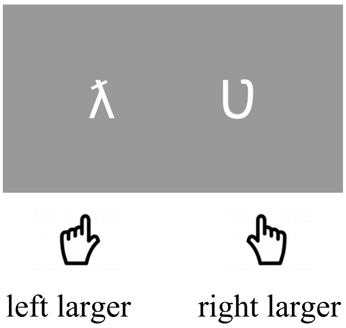


Figure 14. Tasks in the new symbol experiment.

The participants first learned new symbols for the numbers between 1 and 3, and between 7 and 9 (Figure 14). To ensure that the participants have learned them in the learning phase, symbols were practiced until a threshold hit rate was reached. In a trial, a new symbol and an Indo-Arabic digit were shown simultaneously, and the participant decided whether the two symbols denoted the same value by pressing the R or I key.

The stimuli were visible until response. After the response, auditory feedback was given. In a block, all symbols were presented 10 times (60 trials in a block) in a randomized order. In half of the trials, the symbols denoted the same values. The symbol learning phase ended if the error rate in a completed block was smaller than 5% or the participant could not reach that level in five blocks.

In the following comparison task, the participants decided which number is larger in a simultaneously presented new symbol pair by pressing the R or I key (Figure 14). In a trial, two numbers were shown until response, and the participants chose the larger one. Numbers to be compared could be between 1 and 3, and between 7 and 9. After the response, auditory feedback was given. All possible number pairs including the applied numbers, excluding ties, were shown 15 times, thereby resulting in 450 trials.

Presentation of the stimuli and measurement of the responses were managed by the PsychoPy software (Peirce, 2007).

Participants.

Twenty-three university students participated in the experiment for partial course credit. After excluding 4 participants showing higher than 5% error rates (higher than the mean + the standard deviation of the error rates in the original sample) in the comparison task, the data of 19 participants was analyzed (16 females, mean age 22.2 years, standard deviation 4.6 years).

Results

All participants successfully reached a lower than 5% error rate within 3 blocks in the symbol learning task. Therefore, no participants were excluded for not learning the symbols within 5 blocks.

For all participants, the mean error rates and the mean reaction times for correct responses were calculated for all number pairs. Data of participants with higher than 5% mean error rate were excluded (higher than the mean + the standard deviation of the error rates in the original sample). The mean error rates and reaction times of the group are displayed in Figure 15 for the whole stimulus space. Visual inspection of the error rate pattern suggests that partly the value model can be observed, although the data are rather noisy, as reflected in some outlier cells. In the case of reaction time, it is more straightforward that the pattern is more in line with the association model (see the two expected pure patterns in Figure 13). In the reaction time data, one can also observe the end effect: number pairs including the largest number in the range (i.e., 9) are faster to process (Leth-Steensen & Marley, 2000; Scholz & Potts, 1974). (There are different possibilities concerning what causes the end effect. It is possible that participants learn that 9 is the largest number in the actual session; therefore, when 9 is displayed, no further consideration is required in a comparison task. Alternatively, according to the ANS model, it is possible that in the session, number 9 has neighboring number only on one side, and the overlap between the noisy signal distributions should be smaller, thereby leading to a faster response; Balakrishnan & Ashby, 1991).

Number on the right (columns) / **Number on the left** (rows)

	1	2	3	7	8	9
1		4.2%	2.5%	0.0%	0.4%	0.7%
2	3.2%		3.2%	1.4%	0.0%	1.1%
3	0.7%	4.6%		3.5%	1.4%	1.8%
7	0.0%	0.7%	3.2%		5.3%	0.7%
8	0.7%	0.0%	0.7%	4.6%		2.8%
9	0.4%	1.1%	1.1%	0.4%	2.1%	

Number on the right (columns) / **Number on the left** (rows)

	1	2	3	7	8	9
1		1416	1319	1216	1142	860
2	1445		1566	1350	1252	897
3	1387	1698		1532	1227	945
7	1239	1403	1692		1503	957
8	1081	1158	1303	1484		993
9	885	901	889	959	963	

Figure 15. Error rates (left) and reaction times (in ms, right) in the whole stimulus space.

To test the results statistically, we first fit the two predictions of the models (Figure 13) to the group average of the error rate and the reaction time data (Figure 15) with a simple linear regression, where one of the model predictions was the explanatory variable and one of the behavioral performance measurements was the dependent variable. Then the goodness of the fit measured as R^2 was calculated (R^2 columns given in Table 3), and the correlations of two models were compared with the method described by Steiger (1980) for every performance measurement (difference of the group fits column is given in Table 3). As an alternative method, we calculated the R^2 values for every single participant for both the value and association models, and the R^2 of these model fits, as ordinal variables, were compared pairwise with Wilcoxon signed-rank test (Better model for the participants column is given in Table 3).

To fit the distance effect appropriately, the end effect should also be considered, and its variance should be removed from the data. Inspection of the descriptive data on Figure 15 suggests that number pairs including the number 9 were involved in the end effect in the present study. One possibility to remove the end effect is to apply multiple linear regression, and beyond the distance effect regressor, an end effect regressor (e.g., 1 if the number pair includes 9, otherwise 0) also should be utilized. The problem with this solution is that the end effect not only shortened the response latency for number pairs including 9 but it also decreased the slope of the distance effect in those cells (see the less steep distance effect in the row and column with 9 than in other rows and columns). As the end effect is not added linearly to the distance effect, a multiple linear regression could not describe this nonlinear aspect of the end effect, which in turn would distort the distance effect results. As an alternative method, to remove the end effect, all cells with number pairs including 9 were removed from the analysis (i.e., the bottom row and the right column on Figure 15) and only the distance effect regressors were used. Therefore, for all linear fits (Table 3) in both the group average and the participants level, the number pairs including 9 were removed.

Regarding the possible difference between the goodness of fit of the two models, we note that the difference is limited by the fact that the two models correlate, e.g., the value model can be considered as a modified association model with an additional increase of the values in the top-right and bottom-left part of the stimulus space seen in Figure 13. Therefore, if one model is appropriate, then the other inappropriate model should show some non-zero R^2 value too, although the R^2 should be smaller than the R^2 of the appropriate model.

Results for the goodness of fits (Table 3, linear model columns on the left) show that in the error rates, the two models are indistinguishable, and in the reaction time patterns, the association model seems to describe the data better in line with the visual inspection of the data.

Although error rate and reaction time data are highly informative, the recently becoming more popular diffusion model analysis could draw a more sensitive picture (Ratcliff & McKoon, 2008; Smith & Ratcliff, 2004). In the diffusion model, decision is based on a gradual accumulation of evidence offered by perceptual and other systems, and decision is made when appropriate amount of evidence is accumulated. Reaction time and error rates partly depend on the quality of the information (termed the drift rate) upon which the evidence is built. Drift rate is considered to be the most important parameter that influences the number comparison performance and the task difficulty (Dehaene, 2007). Importantly, observed reaction time and error rate parameters can be used to recover the drift rates (Ratcliff & Tuerlinckx, 2002; Wagenmakers et al., 2007).

Drift rates can be more informative than the error rate or the reaction time because drift rates reveal the sensitivity of the background mechanisms more directly (Wagenmakers et al., 2007). To recover the drift rates for all number pairs, the EZ diffusion model was applied (Wagenmakers et al., 2007). The EZ model supposes that some of the parameters do not play a role in the response generation, and the model investigates and recovers only the drift rate, the decision threshold, and the non-decision time parameters. If one can suppose that only these three parameters play a role in the responses, then the EZ model can be utilized. Importantly, one essential advantage of this method is that unlike most other diffusion parameter recovery methods, EZ can be used when the number of trials per cells is relatively small. For edge correction, we used the half trial solution, i.e., for error rates of 0, 50, or 100%, the actual error rate was modified with the percent value of 0.5 trial, e.g., in a cell with 15 trials and 0% error rate, the corrected error rate was 0.5/15, which is 3.33% (see the exact details about edge correction in Wagenmakers et al., 2007). The scaling within-trials variability of drift rate was set to 0.1 in line with the tradition of the diffusion analysis literature. Drift rates for all number pairs and participants were calculated. The mean drift rates of the participants (Figure 16) show a similar pattern observed above for the former descriptive data. Fitting the two predictions of the models, the association model shows again a better fit (Table 3). In addition, (a) the largest difference between the goodness of fit of the two models can be observed for the drift rates (compared to the error rate and the reaction time data) and (b) the highest R^2 value is found for the drift rates, thereby suggesting that the drift rate indeed captures the difficulty of the comparison tasks more sensitively than the error rates or the reaction times do.

Number on the right

Number on the left

	1	2	3	7	8	9
1		0.141	0.156	0.168	0.171	0.208
2	0.144		0.147	0.155	0.171	0.188
3	0.160	0.135		0.134	0.151	0.185
7	0.169	0.155	0.127		0.143	0.183
8	0.189	0.173	0.165	0.140		0.181
9	0.200	0.188	0.197	0.187	0.188	

Figure 16. Drift rate values in the whole stimulus space.

The analysis above supposed that the distance effect (either coming from the value model or from the association model) is linear. However, a logarithmic or a similar function with decreasing change as the distance increases might be a better option to describe the data. First, one cannot suppose a linear distance effect, because after a sufficiently large distance, the reaction time should be unreasonably short or even negative, which would not make sense. Second, in a former artificial symbol comparison task, where the missing size effect did not influence the distance effect, the distance effect was better described with the logarithm function than with a linear function (unpublished results in (Krajcsi et al., 2016)). For these reasons, the analysis of goodness of fit was repeated with logarithmic distance effect models, in which the regressors were the natural logarithm of the values of the previously used linear models seen in Figure 13. The results (Table 3, logarithm model columns on the right) show that (a) for all three data types (error rate, reaction time, and drift rate), the association model fits better than it did with the linear regressor models and (b) the differences of the two models are larger than they were for the linear regressor models. Overall, the largest difference between the value and the association models can be seen in the logarithm model versions for the drift rates.

Table 3. Goodness of fit of the models (measured as R^2) and comparison of the correlations (Difference column) for the error rates, reaction times, and drift rates patterns based on the group average data, and hypothesis tests for choosing the better model based on the participants' data.

	Linear model (Figure 2)				**Logarithm model**			
	Value model R^2	**Association model R^2**	**Difference of the group fits**	**Better model for the participants**	**Value model R^2**	**Association model R^2**	**Difference of the group fits**	**Better model for the participants**
Error rate	0.709	0.708	$Z = 0.008$, $p = 0.993$	$T = 73$, $p = 0.376$	0.714	0.821	$Z = -1.091$, $p = 0.275$	$T = 92$, $p = 0.904$
Reaction time	0.543	0.790	$Z = -2.294$, $p = 0.022$	$T = 44$, $p = 0.040$	0.457	0.817	$Z = -3.646$, $p < 0.001$	$T = 34$, $p = 0.014$
Drift rate	0.526	0.861	$Z = -3.647$, $p < 0.001$	$T = 39$, $p = 0.024$	0.425	0.874	$Z = -5.748$, $p < 0.001$	$T = 18$, $p = 0.002$

While our present main interest is the nature of the distance effect, it is worth to note that no size effect can be found in the data: The regressor formed as the sum of the two numbers to be compared (e.g., the regressor value for the 3 vs. 4 number pairs is 7) does not fit either the error rates (R^2 = 0.001), or the reaction time (R^2 = 0.01), or the drift rate (see below) data (R^2 = 0.001). These data replicate the results of Krajcsi et al. (2016), thereby confirming that in new symbols with equal frequency of numbers in a comparison task, the size effect does not emerge and also confirm that the distance and size effects may dissociate. Relatedly, we note that the size effect could not influence the fit of the distance effect not only because the size effect could not be demonstrated in the present data but also because the size effect regressor (sum of the numbers to be compared) does not correlate with distance effect regressor (difference of the numbers to be compared) at all.

Reliability of the results.

To investigate the reliability of the present results, two additional experiments are summarized here: (a) the whole experiment was repeated with another sample and (b) the data of a follow-up study was analyzed where the same paradigm was used with Indo-Arabic numbers instead of new symbols to see if the distance effect can follow the associations of the numbers and small-large responses in an already well-established notation (Kojouharova & Krajcsi, 2018). (a) In the replication study, 41 university students participated. Four of them were excluded, either because they did not reach the required maximum 5% error rate after 5 blocks of symbol learning or because they used wrong response keys. Five additional participants were excluded, because they had higher than 6.5% error rate (which was the mean + standard deviation error rate in that sample) in the comparison task. As a result, the data of 32 participants were analyzed (mean age was 21.0 years, 3 males). The error rate, reaction time, and drift rate means for the whole stimulus space can be seen in Figure 17, and the R^2s of the models with the appropriate hypothesis tests are displayed in Table 4. While the reaction time and drift rate means replicate the results of the main study (although the difference was significant only with the comparison of the group fits, but not with the hypothesis test choosing the better fit for the participants), the error rates show the superiority of the value model. (b) In the Indo-Arabic comparison task, 23 university students participated. One participant was dyscalculic whose data were excluded from further analysis, and 2 further participants were excluded for having an error rate higher than 5%. Therefore, the data of 20 participants were analyzed (mean age was 20.15 years, 4 males). The goodness of fit of the logarithmic models and their contrast can be seen in

Number on the left \ Number on the right	1	2	3	7	8	9
1		8.5%	2.9%	1.0%	1.0%	0.4%
2	9.2%		5.6%	1.9%	0.4%	0.4%
3	2.3%	8.5%		2.9%	2.1%	1.7%
7	1.0%	1.3%	3.8%		6.9%	2.5%
8	0.4%	0.6%	2.5%	8.5%		8.1%
9	0.6%	0.8%	1.3%	4.0%	9.4%	

Number on the left \ Number on the right	1	2	3	7	8	9
1		1250	1252	1038	970	810
2	1374		1429	1212	1035	845
3	1263	1509		1301	1050	856
7	1118	1270	1386		1245	922
8	985	1047	1136	1194		996
9	823	839	818	967	925	

Number on the left \ Number on the right	1	2	3	7	8	9
1		0.132	0.151	0.175	0.182	0.225
2	0.123		0.132	0.157	0.173	0.205
3	0.154	0.130		0.136	0.170	0.196
7	0.164	0.146	0.134		0.130	0.175
8	0.188	0.173	0.152	0.128		0.146
9	0.206	0.199	0.207	0.161	0.148	

Figure 17. Error rates (top left), reaction times (in ms, top right), and drift rates (bottom) in the whole stimulus space in the replication study.

Table 5. The Indo-Arabic study replicated the results of the main study, and also in the error rates, the association model fitted significantly better than the value model.

Looking strictly at the significance of the results, the replication shows a somewhat different result pattern as the first measurement, because in error rate, the significant differences support the value model instead of the association model, and in reaction time and drift rate, not all hypothesis tests are significant. Clearly, some non-significant effects might reflect not only due to the lack of an effect but also due to the lack of statistical power, and significant effects can also be type-I errors (there is especially a chance for this, when replication studies find opposing significant effects). To evaluate the accumulated data, a mini meta-analysis was run on the three set of data (Maner, 2014). Binary random-effects with the DerSimonian-Laird method (Viechtbauer, 2010; Wallace et al., 2012) was performed on the logarithm model fit data measuring the ratio of participants where the association model was better than the linear model. While the error rate does not show a clear preference for any models (45.9% mean preference for the association model with 95% CI of [17.5, 74.2%]), reaction time and drift rate clearly prefer the association model (76.6% with CI of [65.0, 88.2%] for reaction time and 72.9% with CI [58.2, 87.7%] for drift rate). Taken together, while the reaction time and drift rate show the superiority of the association model, the results of the error rates are ambiguous. It is important to highlight that from the viewpoint of the present question, reaction time and especially drift rates are more relevant. First, reaction time data are usually considered to be more reliable and sensitive than error rate, because error rate and reaction time data measure two strongly correlating constructs. Error rate measures it in a dichotomous scale, whereas reaction time is a continuous scale. Therefore, the latter have more information about the trial performance. Second, drift rate measures the difficulty of the task more sensitively than error rates or reaction times in themselves (Wagenmakers et al., 2007); this is also confirmed by the usually higher R^2 values for drift rates than for reaction times or error rates). Therefore, we consider that reaction times and drift rates reliably reflect the superiority of the association model over the value model. At the same time, it might be a question of future research whether heterogeneous error rates are the result of random noise or whether there are aspects of performance that partly reflects the functioning of the value model.

Table 4. Goodness of fit of the models (measured as R^2) and comparison of the correlations (Difference column) for the error rates, reaction times, and drift rates patterns based on the group average data, and hypothesis tests for choosing the better model based on the participants' data in the replication study.

	Linear model (Figure 2)				**Logarithm model**			
	Value model R^2	**Association model R^2**	**Difference of the group fits**	**Better model for the participants**	**Value model R^2**	**Association model R^2**	**Difference of the group fits**	**Better model for the participants**
Error rate	0.791	0.629	$Z = 2.041$, $p = 0.041$	$T = 130$, $p = 0.012$	0.862	0.724	$Z = 2.081$, $p = 0.037$	$T = 150$, $p = 0.033$
Reaction time	0.610	0.719	$Z = -1.258$, $p = 0.208$	$T = 233$, $p = 0.562$	0.517	0.713	$Z = -2.236$, $p = 0.025$	$T = 196$, $p = 0.204$
Drift rate	0.768	0.914	$Z = -2.727$, $p = 0.006$	$T = 232$, $p = 0.550$	0.695	0.929	$Z = -4.284$, $p < 0.001$	$T = 191$, $p = 0.172$

Table 5. Goodness of fit of the models (measured as R^2) and comparison of the correlations (Difference column) for the error rates, reaction times, and drift rates based on the group average data, and hypothesis tests for choosing the better model based on the participants' data in the Indo-Arabic study (Kojouharova & Krajcsi, 2018).

	Logarithm model			
	Value model R^2	**Association model R^2**	**Difference of the group fits**	**Better model for the participants**
Error rate	0.634	0.825	$Z = -2.766, p = 0.006$	$T = 17, p = 0.001$
Reaction time	0.749	0.917	$Z = -3.737, p < 0.001$	$T = 14, p < 0.001$
Drift rate	0.681	0.864	$Z = -3.080, p = 0.002$	$T = 31, p = 0.006$

To summarize the results, it was found that (a) the association model described the distance effect better than the value model; it measured with reaction time and drift rate, while error rate displayed an inconsistent pattern, (b) drift rate draws more straightforward picture than the reaction time or the error rate data, (c) logarithmic type distance effect describes the data more precisely than the linear distance effect, and finally, (d) size effect is absent in the present paradigm with uniform number frequency distribution.

Discussion

The present work investigated whether the numerical distance effect is rooted in the values of the numbers to be compared or in the association between the numbers and the small-large properties. In a new artificial number notation with omitted numbers, the distance effect measured with reaction time and drift rate did not follow the values of the numbers, as it would have been suggested in the mainstream ANS model (Dehaene, 2007; Moyer & Landauer, 1967) or in the value-based explanation of the DSS model. Instead, the effect reflected the association between the numbers and the small-large categories, as proposed by the association-based explanation of the DSS model or by the delta-rule connectionist model of numerical effects (Verguts et al., 2005). Measured with error rate, the results were not conclusive, so it is the question of additional studies whether the inconsistency in the error rate data is simply noise or there are additional aspects of the distance effect that should be investigated with more sensitive methods.

Together with the present results, several findings converge to the conclusion that the symbolic number comparison task cannot be explained by the ANS. First, unlike the prediction of that model suggesting that distance and size effects are two ways to measure the single ratio effect, symbolic distance and size effects are independent (Krajcsi, 2016), and the distance effect can be present even when no size effect can be observed (shown in the present results and in (Krajcsi et al., 2016). Second, the size effect follows the frequency of the numbers as demonstrated in Krajcsi et al. (2016) and also in the present results, where the uniform frequency of the digits induced no size effect (i.e., the slope of the size effect is zero). Third, the present data demonstrated that the distance effect is not directed by the values of the digits as predicted by the ANS model, but they are influenced by the frequency of the association with the small and large categories (see also the extension of the present findings for Indo-Arabic numbers in (Kojouharova & Krajcsi, 2018).

The present and some previous results also characterize the symbolic numerical comparison task; an alternative model should take the following into consideration: (a) symbolic distance and size effects are independent (Krajcsi, 2016; Krajcsi et al., 2016), (b) the effects are notation independent (the present results and (Krajcsi et al., 2016), (c) the size effect depends on the frequency of the numbers (the present results and (Krajcsi et al., 2016), (d) the distance effect depends on the association between the numbers and the small-large categories (present results), and (e) the distance effect can be described with a logarithm of the difference of the values (present results).

It is again highlighted that these results are not the consequence of the possibility that the new symbols are not related to their intended values and that the independent series of symbols would create a performance pattern similar to the association model prediction, because it was already shown that the new symbols prime the Indo-Arabic numbers, thereby revealing that the new symbols denote their intended values (Krajcsi et al., 2016). The present findings were also replicated with Indo-Arabic numbers (Kojouharova & Krajcsi, 2018).

From a methodological point of view, it is worth to note that in the present comparison task, the drift rate seemed to be the most sensitive index to describe performance, which strengthens the role of the diffusion model analysis, among others in cases when sensitivity and statistical power are essential.

To summarize, the results revealed that in an artificial number notation where some omitted numbers might create a gap, the distance effect followed the association with the small-large properties and not the values of the numbers. This result contradicts the Analog Number System model and the value-based DSS explanation, which suggests that the distance effect is directed by the values or the ratio of the numbers. On the other hand, the result is in line with the alternative association-based DSS explanation and the delta-rule connectionist model, in which the distance effect is directed by the association between the number nodes and the small-large nodes.

Ethics Statement

All studies reported here were carried out in accordance with the recommendations of the Department of Cognitive Psychology ethics committee with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

Author Contributions

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication. Both authors contributed equally to this work.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

We thank Ákos Laczkó and Gábor Lengyel for their comments on an earlier version of the manuscript.

Thesis Study 3
The Indo-Arabic Distance Effect Originates in the Response Statistics of the Task

Petia Kojouharova[1,2] and Attila Krajcsi[2]
[1]Doctoral School of Psychology, ELTE Eötvös Loránd University
[2]Department of Cognitive Psychology, Institute of Psychology, ELTE Eötvös Loránd University

Article published online in *Psychological Research* on 21st July 2018
DOI: https://doi.org/10.1007/s00426-018-1052-1

These authors have contributed equally to this work.

In the number comparison task distance effect (better performance with larger distance between the two numbers) and size effect (better performance with smaller numbers) are used extensively to find the representation underlying numerical cognition. According to the dominant Analog Number System (ANS) explanation, both effects depend on the extent of the overlap between the noisy representations of the two values. An alternative Discrete Semantic System (DSS) account supposes that the distance effect is rooted in the association between the numbers and the "small-large" properties with better performance for numbers with relatively high differences in their strength of association, and that the size effect depends on the everyday frequency of the numbers with smaller numbers being more frequent and thus easier to process. A recent study demonstrated that in a new, artificial digit notation, - where both association and frequency can be arbitrarily manipulated – the distance and size effects change according to the DSS account. Here, we investigate whether the same manipulations modify the distance and size effects in Indo-Arabic notation, for which associations and frequency are already well established. We found that the distance effect depends on the association between the numbers and the "small-large" responses. It was also found that while the distance effect is flexible, the size effect seems to be unaltered, revealing a dissociation between the two effects. This result challenges the ANS view, which

supposes a single mechanism behind the distance and size effects, and supports the DSS account, supposing two independent, statistics-based mechanisms behind the two effects.

Introduction

Source of the numerical distance effect.

Simple number representations are considered to be the foundation of numerical and mathematical cognition. One of the most widely utilized effects to investigate the nature of number representation is the numerical distance effect in a comparison task. When participants compare two numbers in order to decide which one is the larger, performance increases (participants make fewer errors and respond faster) with the increase of the numerical distance between the numbers (Moyer & Landauer, 1967). For example, participants respond faster when they have to select the larger member of the 3-7 number pair (numerical distance equals 4) compared to when they do the same for the 3-4 pair (numerical distance equals 1).

The dominant explanation for the distance effect in a comparison task is that numbers are grounded in the analog number system (ANS) which is a continuous, noisy representation of quantities (Dehaene, 2007). The ANS works according to Weber's law, where the performance depends on the ratio of the numbers to be compared. In this model, the comparison distance effect is considered to be a behavioral consequence of that ratio (Moyer & Landauer, 1967). Thus, according to the ANS model the numerical distance effect is based on the values of the numbers, and more specifically, on the ratio of those values.

There is, however, an alternative plausible explanation for the distance effect. Based on recent studies (Krajcsi, 2016; Krajcsi & Kojouharova, 2017; Krajcsi et al., 2016), it was suggested that a system similar to the mental lexicon or a semantic network, the Discrete Semantic System (DSS) might be responsible for the symbolic numerical effects, including the distance effect. According to the DSS model, numbers might become associated with the “large” and “small” properties during the number comparison task. For example, the number 7 will become relatively strongly associated with the “large” property, while the number 1 will become relatively strongly associated with the “small” property. Two numbers which are further apart will differ to a greater extent in the strength of their associations with the “small” and “large” properties, and consequently, they will be easier to process in the comparison task, thus, resulting in a distance effect. One such implementation of this explanation can be found in a connectionist model by Verguts and his colleagues (Verguts et al., 2005; Verguts & Van Opstal, 2014). In that model a number line layer represents the two numbers to be compared. The number line layer is connected to an output layer, in which “number on the left is larger” or “number on the right is larger” nodes can be found. Responding with “larger” more often will increase the strength (weight) of the association between the number and the “larger” response, and numbers with very different weights will be easier to distinguish than numbers with similar weights. The study found that this association-based model can successfully simulate the distance effect. According to the DSS model and the connectionist model, the distance effect is based on the associations between the numbers and the “large” and “small” properties, and not on the ratio of the values to be compared.[9]

The value-based and the association-based models were contrasted in a recent study (Krajcsi & Kojouharova, 2017). Whereas the ratios of the values and the associations between numbers and the “small-large” properties are strongly correlated in number notations we use in everyday life (such as the Indo-Arabic notation), the two critical properties – values and associations – can be dissociated by using an incomplete set of artificial numbers. The participants learned new symbols for the numbers 1, 2, 3, 7, 8, and 9, and then compared them in a number comparison task. The new number digits were presented with equal probability and it was supposed that the new digits form new associations, which are independent of the already established associations of the well-known Indo-Arabic numbers. In this situation, it can be specified how frequently a digit will be associated with the “larger” or “smaller” properties (see the calculated proportions in Table 6).[10] If the distance effect is based on the values of the artificial numbers, as the ANS predicts, then a distance effect for distance 4 should be observed around the 3-7 gap, e.g., the performance for the 3 vs. 7 values should be similar to any other number pairs with the distance of 4 (see Figure 18). On the other

9 Note that there is an increasing number of works suggesting that the ANS explanation has serious issues in its original form. For example, it has been proposed that symbolic and non-symbolic number processing might rely on different types of representations (Lyons et al., 2015), that symbolic and non-symbolic number representations may have different role in math achievement (Schneider et al., 2017), or that the variance of the performance should also be considered when measuring the ratio effect (Lyons et al., 2015). For more issues see, e.g., in the review of Leibovich and Ansari (2016) or Reynvoet and Sasanguie (2016). However, in the present work we specifically test alternative models in which distance and size effects are independent effects, and in which models it is possible that the distance effect is not originated in the value of the numbers, but in the associations of the numbers.

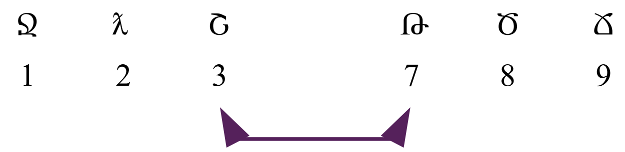
10 The proportion of being smaller or larger is directly proportional to the order of the values. The ordinality of the numbers has been repeatedly proposed to be an important component of number processing. See a recent review of this issue in Lyons, Vogel, and Ansari (2016).

hand, if the distance effect is based on the associations between the numbers and the "small-large" properties, then a distance effect for 1 should be observed, because unlike the value, the associations do not create a gap around the 4-7 pairs (Table 6). For example, performance for the 3-7 number pair should be similar to performance for number pairs with distance 1, such as the 1-2, 2-3 etc. pairs. The results of that study showed that the distance between the artificial numbers followed their associations with the "small-large" responses rather than their values, supporting the association-based DSS account. In other words, it was shown that the distance effect is based on the associations between the numbers and the "small-large" properties.

Table 6. The proportion of being smaller or larger in a number comparison task when the symbols are presented with equal probability.

Example symbols	ᘯ	ƛ	Ꮐ		Ꭹ	Ծ	Ճ
Meaning of the symbols	1	2	3		7	8	9
Proportion of being smaller in a comparison	100%	80%	60%		40%	20%	0%
Proportion of being larger in a comparison	0%	20%	40%		60%	80%	100%

In the study with the new artificial number digits described above, it was supposed that in the case of the well-known Indo-Arabic numbers, the critical properties of the value-based account and the association-based account highly correlate (Krajcsi & Kojouharova, 2017). This is because the numbers have already formed stable associations with the "small-large" properties during years of experience, e.g., 3 and 7 already formed their relatively different associations. However, the supposed rigidity of



distance is 4 (value-based effect) or distance is 1 (association-based effect)

Figure 18. If the numbers between 3 and 7 are omitted, 3 and 7 become neighbors in the sequence. If in a number comparison task, participants decide according to their values, the distance between them should remain 4. If the decision is made based on their association with the "small–large" responses (which is their ordinal position in the sequence), the distance between them should become 1.

the Indo-Arabic notation has not been tested directly. In many cases statistics from the environment can be acquired rather quickly (Dehaene et al., 1993; Fischer, Mills, & Shaki, 2010), and it is possible that the associations between Indo-Arabic digits and "small-large" properties can be reshaped relatively fast. Therefore, one of the main aims of the present study is to investigate whether the Indo-Arabic distance effect could be modified with the use of an incomplete number series, as an association-based distance effect was already demonstrated with new artificial symbols (Krajcsi & Kojouharova, 2017). Additionally, the potential association-based distance effect in Indo-Arabic notation would have the following important implication: The association-based distance effect has been revealed with new artificial digits (Krajcsi & Kojouharova, 2017) and the new digits have been shown to be processed like regular numbers (Krajcsi et al., 2016) (demonstrating the priming distance effect between the new symbols and the Indo-Arabic numbers), but still, the new notation could be processed differently in many – yet unspecified – ways, making the new-symbol based evidence for the association-based model less convincing. Therefore, if an association-based distance effect could be observed for Indo-Arabic digits, it would mean a more direct evidence in support of the association-based explanation.

Source of the numerical size effect.

The numerical size effect is another simple effect that is used to investigate the nature of numerical representations. In a number comparison task, performance decreases (response time and error rate both increase) as the value, i.e., the numerical size of the numbers increases, hence the term size effect.

According to the widespread explanation, the ANS could be responsible for this effect, too. The ANS account proposes that the size effect is another consequence of the ratio-based performance: Larger numbers with the same distance form smaller ratios, and the relatively bad performance with larger numbers is the consequence of the small ratio. For example, performance is better for the 1-2 number pair (ratio is 2) than for the 8-9 number pair (ratio is 1.125).

As an alternative explanation, the DSS model proposes that the size effect is in fact a frequency effect. Smaller numbers are used in everyday life more frequently than larger numbers, and analysis of various corpora revealed that the frequency of a number is proportional to the power of that value (Dehaene & Mehler, 1992). Because it has been demonstrated repeatedly that relatively rare stimuli are harder to process, it is possible that for this reason larger numbers are slower and more error-prone to process, which could also explain the size effect in comparison tasks. To test this explanation, in a recent study, participants learned new symbols and compared them, while the frequency of the numbers was manipulated (Krajcsi et al., 2016). It was found that the size effect followed the frequency of the new symbols, e.g., if all symbols were presented with equal frequencies, then the size effect did not appear.

In the previously described study new symbols were utilized to test the size effect, partly because it was supposed that the statistical knowledge about the over-learned Indo-Arabic numbers is hard to modify. Additionally, the size effect in Indo-Arabic numbers has been demonstrated in several studies, and in most studies equal frequencies of the number digits were used (Moyer & Landauer, 1967), suggesting that the Indo-Arabic size effect cannot be modified by manipulating the frequencies of the numbers in a single session. However, to our knowledge it has not been investigated whether the size effect begins to adapt to the statistics of the session, i.e., whether the size of the size effect decreases throughout the session if participants are exposed to uniform digit distribution. The second main aim of this study is to investigate whether the size effect decreases when the Indo-Arabic numbers are presented with equal probability.

Aims of the study.

Whereas it has been demonstrated that the DSS model offers a better explanation in a number comparison task with new symbols, the potential plasticity of the distance and size effects has not been investigated in the well-known Indo-Arabic notation. The aim of the present study is to examine whether the distance and size effects can be modified in an Indo-Arabic number comparison task, when the statistics of the actual session deviates from the everyday statistics. More specifically, first we investigate whether the distance effect can be modified when the strength of associations between the “small-large” properties and the numbers is modified by omitting numbers 4, 5, and 6. Second, we investigate whether the size effect gradually decreases throughout the session when the digits are shown with equal frequencies. If any of these modifications occurs, this would strengthen the role of the DSS in symbolic number comparison, and would question the role of the ANS in symbolic number processing.

Methods

The participants compared the Indo-Arabic numbers 1, 2, 3, 7, 8, and 9 by selecting the larger member of a number pair.

Participants.

Twenty-three university students participated in the experiment for partial course credit. All but one were right-handed, and all had normal or corrected to normal vision. One participant was dyscalculic whose data were excluded from further analysis. Two further participants were excluded for having an error rate higher than 5%. Thus, the data of twenty participants were analyzed (16 females, mean age *M* = 20.15 years, *SD* = 2.28 years). The present study was carried out in accordance with the recommendations of the Department of Cognitive Psychology ethics committee. All participants gave written informed consent in accordance with the Declaration of Helsinki.

Stimuli and procedure.

Indo-Arabic numbers were presented in pairs to the participants in a number comparison task. On each trial, the two numbers remained visible until response, and the participants had to choose the larger number by pressing the R and I keys of a computer keyboard (Figure 19). The pairs consisted of all possible combinations of the numbers 1, 2, 3, 7, 8, and 9, excluding ties. The numbers were presented in three blocks, and each number pair was presented 10 times in a block, resulting in a total of 900 trials (300 trials per block). The participants could rest briefly between the blocks. Each number was presented with equal frequency.

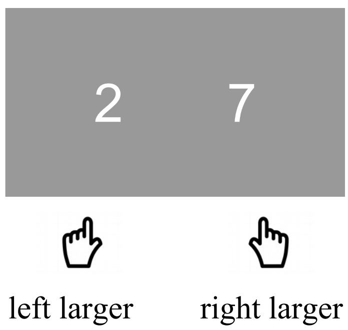


Figure 19. Number comparison task in the experiment. A pair of numbers was presented to the participants who decided which of the numbers (number on the left or number on the right) is larger by pressing a key.

The presentation of the stimuli and recording of the responses were managed by the PsychoPy software (Peirce, 2007).

Data analysis.

Generally, the data analyses investigate whether it is the value-based or the association-based explanation that describes the data better. To contrast the two descriptions, linear regression is utilized. Instead of the usual distance and size effect calculations (i.e., mean error rates or reaction times are calculated for all distance and size values, then the slopes for the effects are calculated), all effects are investigated and tested in the whole stimulus space, as e.g., in Figure 20. The figure shows performance or performance prediction for each number pair presented during the experiment with columns and rows denoting the two numbers to be compared (all possible pairs, excluding ties), and cells containing expected performance. This method is more informative than the usual distance and size effect calculations, because any systematic deviation from the expected patterns of the effects could be observed. Additionally, with the relatively large number of cells the presence of any systematic pattern could be convincing even without the statistical hypotheses tests.

Quantifying the distance and the size effects in the whole stimulus space.

First, we describe how the distance and size effects were quantified for the whole stimulus space. These effects will be later used as regressors in the linear regression.

The top panel of Figure 20 depicts the distance effect as predicted by the value-based and the association-based accounts. The distance effect appears as an improvement in performance from the top-left and bottom-right diagonal towards the bottom-left and top-right corners of the stimulus space. For the value-based model the cell values were calculated as the logarithm of the absolute value of the difference of the two numbers (i.e., logarithm of the difference of the pairs comprised of 1, 2, 3, 7, 8, and 9), whereas for the association-based model they were computed as the logarithm of the distance of their order (i.e., logarithm of the difference of the pairs comprised of 1, 2, 3, 4, 5, and 6) (see also Footnote 10). Logarithm of the distances instead of the linear distances was used, because (a) linear distance effect would cause negative error rate and reaction time values when the distance is very large, which values would not make sense, and (b) previous data showing purely the distance effect without the size effect suggested that the distance effect can be described more appropriately with the logarithm function instead of the linear version (Krajcsi & Kojouharova, 2017; Krajcsi et al., 2016) (See the possibilities for alternative formulations of the distance and size effects in the Results section.) The cell values of those performance predictions were later used as distance effect regressors in a multiple linear regression model fitting (see details below). As can be observed, the two explanations predict different performance not only for the two numbers next to the gap (3 and 7), but also for all number pairs whose members are on opposite sides of the gap, e.g., the distance between 1 and 9 is 8 in the value-based, and 5 in the association-based model.

The bottom panel of Figure 20 shows the size effect regressor computed as the sum of the two numbers of each pair. The size effect appears as a decrease in predicted performance from the top-left towards the bottom-right corner.

Figure 21 portrays a possible linear combination of the distance and size effects, illustrating what type of performance pattern can be expected. Additionally, if the size effect was absent as the consequence of the uniform distribution of the digits in the session, then the results should resemble the predicted performance for the distance effect (top panel in Figure 20).

Calculating the error rate, the reaction time, and the drift rate.

In the present analysis, first, we calculated (a) the mean error rates, (b) the mean reaction times of the correct responses, and (c) the drift rates for each participant and for the whole stimulus space (i.e., for each presented number pair) (see more details about the drift rate below). In the case of reaction time, extreme values above 2000 ms were excluded which resulted in the removal of a total of 43 trials (0.23% of all trials).

As already mentioned, for all participants and all number pairs, the drift rates were calculated. Drift rate is a part of the increasingly popular diffusion model analysis, and is assumed to provide a more sensitive measure of performance (Ratcliff & McKoon, 2008; Smith & Ratcliff, 2004), as it combines various properties of both the error responses and the reaction time distribution. In this model evidence is accumulated gradually from perceptual and other systems until a sufficient amount of evidence becomes available for a decision to be made. Drift rate represents the quality of information upon which the evidence is built, and while error rates and reaction times adequately capture performance on a task, drift rate is more directly related to the background mechanisms of performance. Furthermore, drift rates can be recovered

Distance effect

Value-based model

Number 1 \ Number 2	1	2	3	7	8	9
1		0.00	0.69	1.79	1.95	2.08
2	0.00		0.00	1.61	1.79	1.95
3	0.69	0.00		1.39	1.61	1.79
7	1.79	1.61	1.39		0.00	0.69
8	1.95	1.79	1.61	0.00		0.00
9	2.08	1.95	1.79	0.69	0.00	

Association-based model

Number 1 \ Number 2	1	2	3	7	8	9
1		0.00	0.69	1.10	1.39	1.61
2	0.00		0.00	0.69	1.10	1.39
3	0.69	0.00		0.00	0.69	1.10
7	1.10	0.69	0.00		0.00	0.69
8	1.39	1.10	0.69	0.00		0.00
9	1.61	1.39	1.10	0.69	0.00	

Size effect

Number 1 \ Number 2	1	2	3	7	8	9
1		3	4	8	9	10
2	3		5	9	10	11
3	4	5		10	11	12
7	8	9	10		15	16
8	9	10	11	15		17
9	10	11	12	16	17	

Figure 20. Models used in the analysis. Top. The distance effect regressors for the stimulus space used in the present study based on the value-based model (left side) and on the association-based model (right side) are shown in the top panel. The regressors were calculated as log(large− small), where log is natural logarithm, large and small are the large and small numbers of the pair, and either the value of that number (in the value-based model) or the order of that number (in the association-based model) was used (e.g., 7 is in the 4th place of the ordered set). Bottom. The size effect regressor which was computed as large+small.

based on observed error rate and reaction time parameters (Ratcliff & Tuerlinckx, 2002; Wagenmakers et al., 2007). Here, we applied the EZ-diffusion model (Wagenmakers et al., 2007), a simplified version of the diffusion model which still allows for the recovery of drift rates in the case of sparse data from a relatively small number of parameters. For edge correction we used the half-trial solution (see the exact details about edge correction in Wagenmakers et al., 2007). The scaling within-trials variability of drift rate was set to 0.1 in line with the tradition of the diffusion analysis literature. See additional technical explanations about the EZ-diffusion model in Wagenmakers et al. (2007).

Distance and size effects

Value-based model

Number 1 \ Number 2	1	2	3	7	8	9
1		0.30	-0.29	-0.99	-1.05	-1.08
2	0.30		0.50	-0.71	-0.79	-0.85
3	-0.29	0.50		-0.39	-0.51	-0.59
7	-0.99	-0.71	-0.39		1.50	0.91
8	-1.05	-0.79	-0.51	1.50		1.70
9	-1.08	-0.85	-0.59	0.91	1.70	

Association-based model

Number 1 \ Number 2	1	2	3	7	8	9
1		0.30	-0.29	-0.30	-0.49	-0.61
2	0.30		0.50	0.21	-0.10	-0.29
3	-0.29	0.50		1.00	0.41	0.10
7	-0.30	0.21	1.00		1.50	0.91
8	-0.49	-0.10	0.41	1.50		1.70
9	-0.61	-0.29	0.10	0.91	1.70	

Figure 21. Possible combination of the distance and size effects according to the models, showing the patterns of possible behavioral performance. The combined pattern was calculated according to the a_1 *log(large −small)+*a_2 *(large +small)+*b *formula. Parameter* a_1 *is set to − 1 to align the direction of the change for the two effects,* a_2 *is set to 0.1 for scaling purposes, and* b *is set to 0. The specific values of the parameters are arbitrary with the constraint that a1 should be negative, and the relative weights of the distance and size effects should follow the effect sizes of those effects in the behavioral patterns. Darker shade indicates worse performance.*

Analysis of the distance effect.

To investigate the distance effect, the value-based and the association-based distance effects were contrasted in error rates, reaction times, and drift rates. Five analyses were utilized. (1) In a group level analysis, two multiple linear regression analyses were performed: first, the value-based distance (logarithm of the distance of the two values) and the size effects (sum of the two values) were fitted (top-left and bottom regressors in Figure 20) to the group average data (top row in Figure 22), second, the association-based distance (logarithm of the distance of the orders) and the size effects (sum of the values) were fitted (top-right and bottom regressors in Figure 3) to the group average data (top row in Figure 22). Finally, to contrast the two models, the R^2 of those two fittings were compared.[11] Note that in this analysis the two competing models are not included in the same regression as two regressors, but the two models are investigated in two different regressions and the measurements of the fit are

11 An additional effect which may appear in a number comparison task is the end effect—reaction time is much faster and error rate is lower for the cells containing the largest number of the set, in this case all cells with number 9. Visual inspection suggests that there might be an end effect in the data. As this effect can distort the stimulus space in a non-linear manner (it flattens the distance effect in the relevant cells), an additional analysis was performed with the cells containing the number 9 excluded. The results were similar to the ones described in the text.

compared. It is also important to note that the distance effect regressors for the value-based and the association-based models highly correlate ($r = 0.86$, $p < 0.001$) because of the structure of the stimulus space. Thus, if one of the models is appropriate, the other models will also be appropriate, albeit with a smaller R^2 value. (2) In an individual level analysis, the same multiple linear regressions were run on individual data, i.e., the regression was repeated for all participants. Then the R^2 of the participants for both models were compared with a non-parametric paired samples test (Wilcoxon signed-rank test as R^2 values can be considered to be an ordinal variable). (3) The same analysis as described in the previous point (analysis 2) was repeated, with the trials divided into three blocks. This allowed us to examine whether there is a slow transition from one model to the other with the progress of the trials. R^2 values were calculated for the error rate, reaction time, and drift rate, for both models, as in the individual analysis, however, they were not calculated on the whole session, but separately for the three blocks. Then, a 3×2 repeated measures analyses of variance were run with individual R^2 as the dependent variable and with the block (Block 1, 2, and 3) and the model (value-based and association-based) as factors, where an interaction between the factors would indicate a transition between the models. (4) In the previous methods, the two models (i.e., two types of distance effect regressors) were used in separate regressions, because the two models strongly correlate. However, in a following hierarchical multiple regression analysis, the three regressors were applied at the same time: The value-based distance effect, the association-based distance effect, and the size effect. In the first block, either the size and value-based distance or the size and the association-based distance regressors were applied, while in the second block the other distance (either the association- or the value-based) regressor was added, and it was investigated whether the newly added distance regressor can improve the fit of the whole model. The regression analysis was run for the group average error rate, the reaction time, and the drift rate data. (5) In a last analysis, a single multiple linear regression was performed with the three regressors at the same time, but here the regression was run for all participants' data, and not for the group average data. After performing the regression, it was tested whether the weights (slopes) of the two distance regressors deviate from 0.

Analysis of the size effect.

To investigate the size effect, two analyses were utilized. (1) The slope of the size effect regressor was taken from the multiple linear regression analysis described above for the participant whole session data (analysis 2) for error rate, reaction time, and drift rate for each participant, then a one-sample *t*-test was performed to determine whether it deviated significantly from 0. (2) Additionally, to test whether the size effect decreased across the session, the change in the slope of the size effect regressor (taken from the three blocks analyses described above, analysis 3) across the three blocks of the experiment was compared as repeated measurements.

Results

Mean error rates, mean reaction times for the correct responses, and drift rates were calculated for each participant overall and also for each block, then averaged across participants for the whole stimulus space. The results are presented in Figure 22. Visual inspection of the data suggests an advantage for the association-based model, although less certain in the case of reaction time. The presence of the size effect is also visible.

Type of the distance effect.

A multiple linear regression analysis was conducted on the averaged data of the group for error rates, reaction times, and drift rates with distance (for each model) and size as regressors (see Figure 20 for the regressors) (distance effect analysis 1, left panel of Figure 23). The same analysis was performed for each participant separately, and the goodness of fit (R^2) values were entered as ordinal variables in a Wilcoxon signed-rank test for a pairwise comparison (see the Methods section for more details). The results are summarized in the right panel of Figure 23 (distance effect analysis 2).

The association-based model shows a better fit than the value-based model in the case of error rates, reaction times, and drift rates. Moreover, this difference was statistically significant for all indexes.

A hierarchical multiple regression analysis (distance effect analysis 4) also confirmed the previous results. Adding the association-based distance regressor to the value-based distance and size regressors significantly increased the R^2 value of the model ($F(1, 26) = 30.342$, $p < 0.001$ for error rates, $F(1, 26) = 57.593$, $p < 0.001$ for reaction time, and $F(1, 26) = 35.420$, $p < 0.001$ for drift rates), while adding the value-based distance regressor to the association-based distance and size regressors did not increase the R^2 of the model ($F(1, 26) = 1.007$, $p = 0.325$ for error rates, $F(1, 26) = 1.599$, $p = 0.217$ for reaction time and $F(1, 26) = 0.163$, $p = 0.689$ for drift rates).

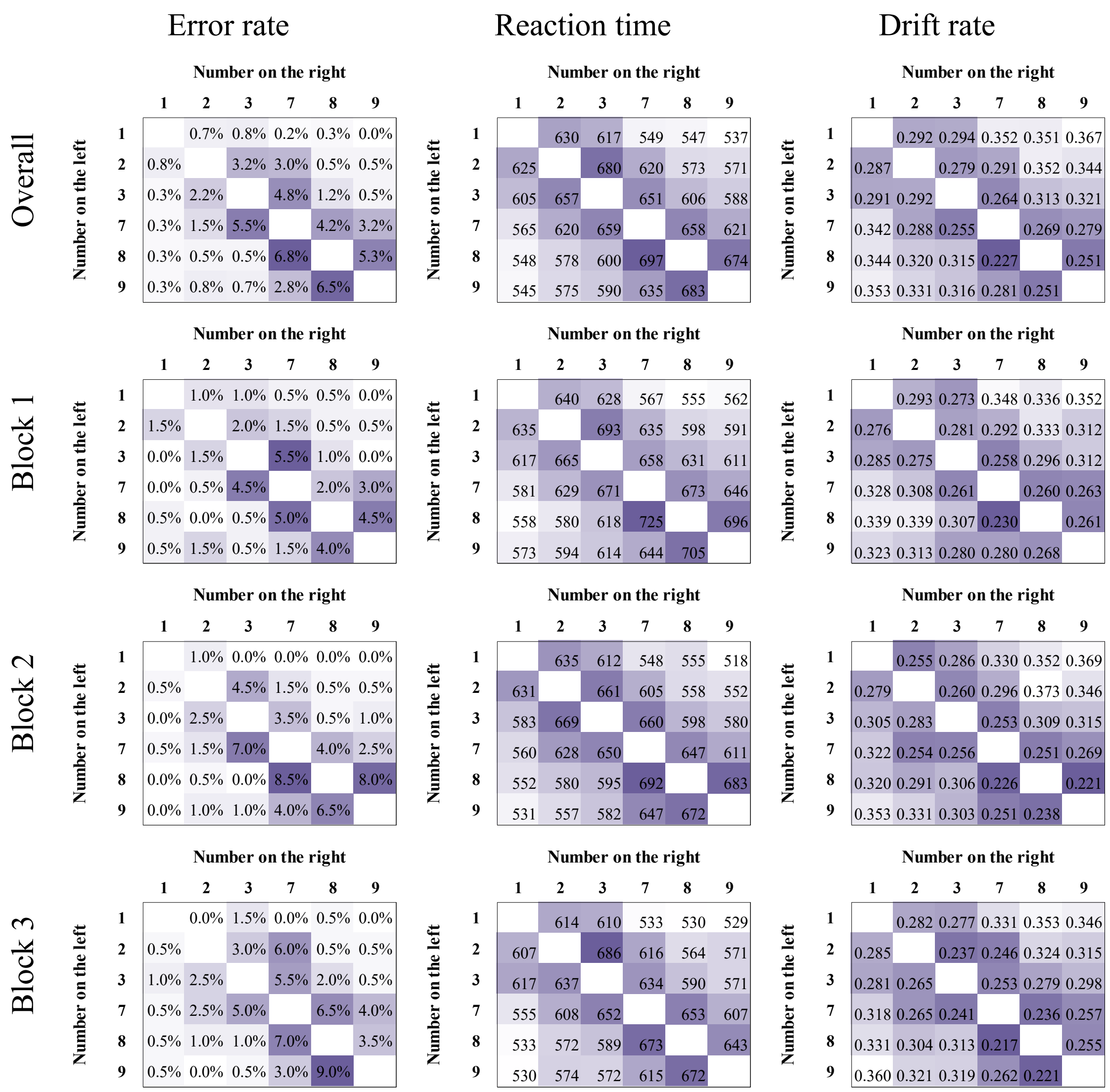


Figure 22. *Error rate, reaction time (in ms), and drift rate values for the whole stimulus space for the whole session and for the three blocks. Darker shade indicates a decrease in performance.*

A similar analysis performed on the individual data (distance effect analysis 5) also found that while the weight of the association-based distance effect deviates from zero (M = 3.2%, $t(19) = -5.86$, $p < 0.001$; M = -66.0 ms, $T = 0$, $p < 0.001$; $M = 0.054$, $T = 3$, $p < 0.001$, for the error rates, reaction time, and drift rate, respectively), the value-based distance effect does not contribute to the variance of the performance (M = 0.4%, $t(19) = 1.37$, $p = 0.185$; M = -7.7 ms, $t(19) = -1.48$, $p = 0.155$, $M = 0.003$, $t(19) = 0.556$, $p = 0.585$). These results mean that when the two distance effect are implied in the same regression, it is only the association-based distance effect that contributes to the variance of the performance, and in regressions with a single distance effect regressor, the value-based distance effect predicts the performance only because it correlates with

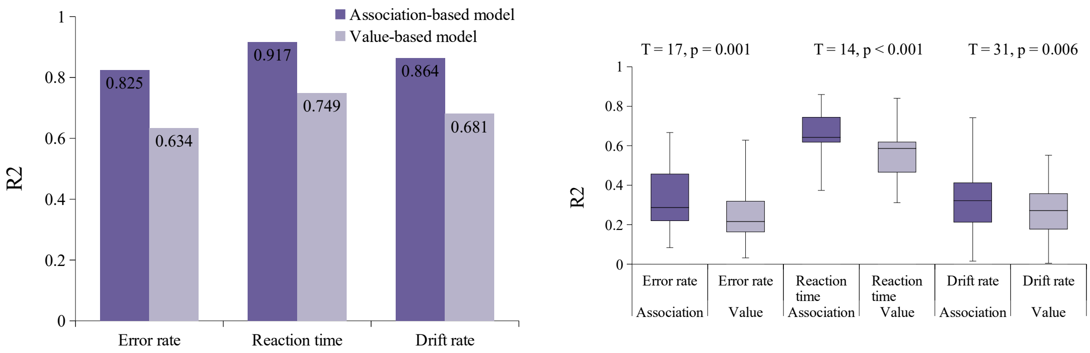


Figure 23. Left: goodness of fit of the models (measured as R^2) for the error rates, reaction times, and drift rates based on the group average data (distance effect analysis 1). Right: box plot of the goodness of fit and hypothesis tests of the difference of the fit for the value-based and association-based models based on the participants data (distance effect analysis 2).

the association-based distance effect.

A 3×2 repeated measures analysis of variance was conducted on the R^2 values with the factors block (Block 1, 2, and 3) and model (value-based and association-based) in order to account for a potential transition between the models which may have occurred during the task (R^2 values are summarized in Table 7) (distance effect analysis 3). Only a main effect of the model was found in the case of error rates ($F(1,17) = 38.858$, $p < 0.001$, $\eta_p^2 = 0.696$), reaction times ($F(1,19) = 15.133$, $p = 0.001$, $\eta_p^2 = 0.443$), and drift rates ($F(1,19) = 14.058$, $p = 0.001$, $\eta_p^2 = 0.425$). There was no main effect of block (all *F*s between 0.413 and 2.018, *p*s = 0.147-0.665), and more importantly, there was no interaction (all *F*s between 0.382 and 1.474, *p*s=0.243-0.685). Thus, not only was the association-based model a better fit overall, but it also described the participants' data better from the very beginning of the session.

Table 7. Goodness of fit (R^2) for error rate, reaction time, and drift rate for the three blocks of the experiment for the group average data (distance effect analysis 3).

	Block 1		Block 2		Block 3	
	Value-based	Association-based	Value-based	Association-based	Value-based	Association-based
Error rate	0.452	0.682	0.657	0.779	0.528	0.696
Reaction time	0.756	0.895	0.734	0.925	0.689	0.846
Drift rate	0.633	0.739	0.664	0.813	0.621	0.810

To display the distance effect in a way that is more in line with the methods used in the literature, the distance effect was also calculated as mean errors or mean reaction times for all distance values and for all participants, then the group mean was calculated. Top of Figure 24 on left shows the distance values for the number pairs (the logarithm of these values can be seen in Figure 20), and cells with the same distance values were collapsed. Practically, the classic distance effect calculation utilizes the value-based model. Error rate and reaction time as a function of the distance can be seen in the right panel of Figure 24. It is visible that the distance effect does not follow the usual curve, but there is a discontinuity between the distance 2 and 4, which is the consequence of the distortion caused by the unusual association pattern formed by the 4-6 gap in the present paradigm. More specifically, the distance values come from the value-based model, which handles the large distance values incorrectly, as demonstrated in the previous analyses, and in fact, the distance values 4-8 should be handled as 1-5 (see also Figure 20). This is also the reason why the performance for 4-5 distances shows similar error rate and reaction time as the 1-2 distance cells. On the other hand, if distance is calculated based on the association strength (or in other terms, based on the order of the numbers) as seen at the bottom of Figure 24, the usual distance effect can be observed, which again demonstrates that the distance effect is rooted in the association of the numbers and not in their value.

Size effect.

To assess the presence of the size effect, the slope of the size effect regressor was tested against 0 (size effect analysis 1). As the size effect regressor (sum of the numbers to be compared) does not correlate with distance effect regressor (difference of the numbers to be compared) in either of the two models, its contribution to the model, and thus its slope, should remain the same, independent of which of the two models is fitted. The results revealed that the slope of the size effect was significantly different from 0 for error rates (M = 0.003, 95% CI [0.002, 0.004], $t(19) = 5.45$, $p < 0.001$, d = 1.250), reaction times (M = 2.498, 95% CI [1.096, 3.900], $t(19) = 3.73$, $p = 0.001$, d = 0.856), and drift rates (M = -0.002, 95% CI [-0.004, -0.001], $t(19) = -4.16$, $p < 0.001$, d = -0.954). Figure 25 summarizes the average slopes and confidence intervals for the size effect regressor for each of the three blocks (size effect analysis 2). Although the descriptive data indicate an increase in the slope for error rate and a decrease in the

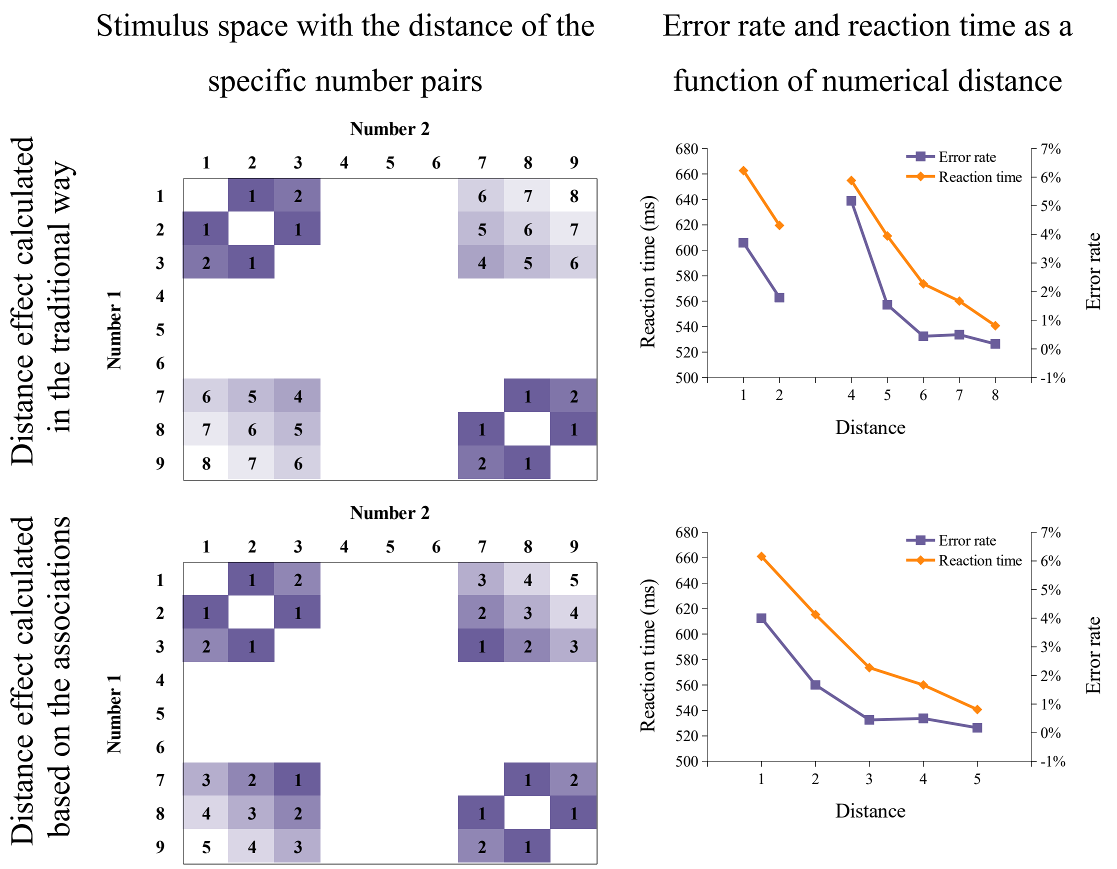


Figure 24. *Distance effect calculated in the traditional way (top) and based on the associations (bottom). The figure shows the whole stimulus space with the distance of the specific number pairs (left), and the error rate and reaction time as a function of numerical distance (right).*

slope for reaction time, none of these changes were significant with *p*s between 0.182-1.000 for the Friedman test.

Sign of learning effects.

One might suggest that in the present data no direct learning can be observed, because there are no performance changes between the blocks. However, we must suppose that the initial state (before the experiment) should be either a value-based model predicted distance effect (prediction of the ANS model, and the prediction of the DSS model in case of learned statistics from former experience), or a “blank” performance (i.e., the performance does not depend on the distance). The latter can be a prediction of the DSS model, if the distance effect works only within a session, and former experience would not influence the actual performance. Because the observed pattern in the first block already deviates from these possible initial states, we can only conclude that learning has occurred.

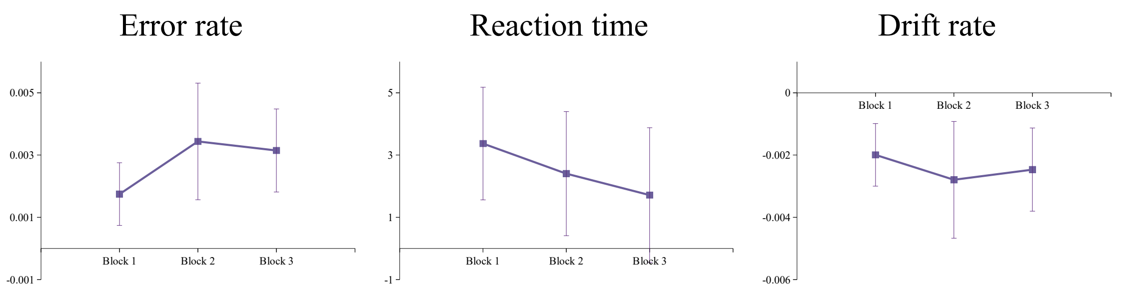


Figure 25. *Means of the slope of the size effect for error rates, reaction times, and drift rate for the three blocks of the experiment. Error bars indicate the 95% confidence intervals.*

Reliability of the distance and size effects.

Distance and size effect slopes are often utilized as potential predictors of other numerical performance, such as math grades in school or general math performance (Goffin & Ansari, 2016; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013). For that purpose it is not only the supporting mechanism that is important in such studies (which is the main question of the present study), but also the reliability of those indexes. To test the reliability of the reaction time effect slopes in the present data, the test-retest correlations were calculated between the first and second blocks and between the second and third blocks for the association-based distance effect, the value-based distance effect, and the size effect. While both the association-based distance effect (r(18) = 0.738, 95% CI [0.439, 0.890], $p < 0.001$ and r(18) = 0.583, 95% CI [0.190, 0.815], $p = 0.007$, for the block 1-block 2 and for the block 2-block 3 correlations, respectively) and the value-based distance effect (r(18) = 0.757, 95% CI [0.473, 0.899], $p < 0.001$ and r(18) = 0.624, 95% [0.251, 0.836], $p = 0.003$) displayed relatively high test-retest correlations, the size effect was not reliable (r(18) = 0.164, 95% CI [-0.300, 0.566], $p = 0.489$, and r(18) = 0.354, 95% CI [-0.105, 0.689], $p = 0.125$). The main reason for the low reliability of the size effect can be the relatively low effect size compared to the effect size of the distance effect, while the measurement noise in those effects could be approximately the same (see the descriptive data to find the different slopes across the size and distance effects on Figure 22). Still, the distance effect was found to be reliable which makes it an appropriate index in correlational studies, although its meaning should be reconsidered according to the present results.

Alternative functions for the distance and size effects.

In the present analysis, the distance effect in the model was formulated as the logarithm of the distance, and the size effect was formulated as the sum of the two values. However, one could imagine that other functions would describe these phenomena more precisely. While there could be debates about the exact formulation of the distance and size effects functions, it most probably would not modify our results. In the case of the distance effect, it is no matter how the distance (or even ratio) effect is formulated, because value-based models would predict a gap between 3 and 7 (see top right model in Figure 20), independent of the exact shape caused by the specific function. Importantly, our data do not reveal a gap between 3 and 7 (Figure 22), which result cannot be described in any alternative formulation. In the case of the size effect, performance change with increasing numbers could be detected with most reasonable size effect formulations, and based on the descriptive data (Figure 22) it is hard to suppose that a large enough slope change that could be significant could be detected. (The group mean results with the model fitting can be downloaded from https://osf.io/qjymb/ where alternative functions can be tested interactively.)

To summarize, according to our results, (1) the association-based model described the participants' data better than the value-based model, (2) the association-based model was dominant from the very beginning of the session with no significant change across trials, (3) the size effect was present for the whole duration of the task, again, with no significant change across blocks.

Discussion

In the present study we investigated the plasticity of the distance and size effects in the Indo-Arabic notation.

First, we tested whether the numerical distance effect in Indo-Arabic notation depends on their value as suggested by the ANS or stems from their associations with the "small-large" properties in accordance with the DSS account. For that purpose only the numbers 1, 2, 3, 7, 8, and 9 were presented in a number comparison task. The results showed that the association-based model explains the data better than the value-based model, and this advantage is observable from the very beginning of the session. The same results have been seen in a new artificial number notation (Krajcsi & Kojouharova, 2017), and now have now been demonstrated in the Indo-Arabic numbers. The present result means that the distance effect depends on the association between the numbers and the “small-large” responses, and this effect is very flexible even in the overpracticed Indo-Arabic number notation. Also, because this result replicates similar findings with new artificial symbolic notation, it confirms that new symbols and Indo-Arabic notation are processed in a similar way.

The second goal of the experiment was to examine whether the size effect could be modified gradually if the numbers were shown to the participants with equal frequency. The results suggest that the size effect remains non-zero in the number comparison task, and even if there is a slight decrease in the size effect (as suggested by the descriptive data), this effect size is too small to be significant in the present sample. It shows that in the already well-known Indo-Arabic notation, the size effect cannot be changed entirely, and it is less flexible than the distance effect.

The present results challenge the ANS model in several ways. First, the distance effect does not depend on the values of the numbers, but on the associations of the numbers with the “smaller-larger” responses. Second, the distance and size effects are independent of each other, thus they cannot be the two consequences of the same ratio effect (Weber’s law) as suggested by the ANS model. The independence of the two effects was also demonstrated in a correlational study, where the slopes of the distance and size effects of a symbolic comparison did not correlate (Krajcsi, 2016), and in a new, artificial digit comparison task where the distance effect was observable even when the size effect was absent (Krajcsi et al., 2016). Similar to that correlational study, the present data also confirm the low correlation of the association-based distance and the size effect slopes in the reaction time: $r_s(18) = 0.347$, $p = 0.133$, 95% CI [-0.112, 0.685].

While the present results challenge the ANS model, these findings are in line with the DSS model. It is possible that in a discrete network, the distance effect is rooted in the associations of the number nodes and the “small”-“large” nodes (Krajcsi et al., 2016; Verguts et al., 2005; Verguts & Van Opstal, 2014). According to the DSS account, the size effect may be rooted in the frequency of the stimuli, and because the distance and size effects stem from two different mechanisms, the two effects may dissociate, and their flexibility may be different.

To summarize, the present study demonstrated that in the case of Indo-Arabic numbers the distance effect in a number comparison task originates in the strength of the associations between the numbers and the “small-large” properties, and these associations are subject to rapid changes depending on the statistics in a session.

Furthermore, the size effect was shown to be relatively unaltered, thus dissociating from the distance effect. The results strengthen the argument that symbolic number processing cannot be explained by the ANS model, instead the results are in line with the DSS explanation.

Compliance with Ethical Standards

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

Conflict of Interest: The authors declare that they have no conflict of interest.

All procedures performed in studies involving human participants were in accordance with the ethical standards of the Department of Cognitive Psychology ethics committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent was obtained from all individual participants included in the study.

Acknowledgments

We thank Krisztián Kasos and Ákos Laczkó for their comments on an earlier version of the manuscript.

Thesis Study 4
Two Components of the Indo-Arabic Numerical Size Effect

Petia Kojouharova (Петя Кожухарова)[1,2] and Attila Krajcsi[2]
[1]Doctoral School of Psychology, ELTE Eötvös Loránd University, Hungary
[2]Department of Cognitive Psychology, Institute of Psychology, ELTE Eötvös Loránd University, Hungary

Article published online in *Acta Psychologica* on 5th of December, 2018
DOI: https://doi.org/10.1016/j.actpsy.2018.11.009

These authors have contributed equally to this work.

In the symbolic number comparison task, the size effect (better performance for small than for large numbers) is usually interpreted as the result of the more general ratio effect, in line with Weber's law. In alternative models, the size effect might be a result of stimulus frequency: smaller numbers are more frequent, and more frequent stimuli are easier to process. It has been demonstrated earlier, that in artificial new number digits, the size effect reflects the frequencies of those digits. In the present work we investigate whether frequency also directs the size effect in Indo-Arabic numbers, in which notation, unlike in new symbols, the frequencies are already firmly established for the participants. We found that frequency has an effect on the size effect in Indo-Arabic notation, but this influence is limited. However, this limited size effect change is acquired fast at the beginning of the session. We argue that these results are more in line with the frequency-based accounts of the size effect.

Highlights

- The numerical size effect follows in part the frequency of Indo-Arabic numbers.
- The change in frequency is incorporated early, but then remains stable.
- There is also a stable component to the numerical size effect.
- The frequency-based account is a better explanation for the numerical size effect.

The Source of the Size Effect

In the field of mathematical cognition, numerical distance and size effects are among the most intensively investigated phenomena. In a number comparison task, the distance effect refers to better performance (e.g., lower error rates or faster responses) when the distance between the two numbers to be compared is relatively large (supposing the size of the stimuli is controlled for), and the size effect refers to better performance when the numbers to be compared are relatively small (supposing the distance between the numbers is controlled for) (Moyer & Landauer, 1967). For example, according to the distance effect it is easier to compare the 1-9 pair (distance 8) than the 3-7 pair (distance 4), and according to the size effect it is easier to compare the 1-3 pair (average size 2) than the 7-9 pair (average size 8).

The mainstream literature assumes that both distance and size effects result from the ratio effect, i.e., the performance depends on the ratio of the two values, in line with Weber's law. The ratio effect suggests that numbers are processed by an evolutionary ancient and simple representation, the Analogue Number System (ANS), similar to the representation processing simple physical properties (Dehaene, 1992, 2007; Moyer & Landauer, 1967).

Alternatively, some other models suggest that the size effect is a frequency effect. In everyday life, small numbers are more frequent than large numbers, and the frequency of the numbers is proportional to the power of the values (Dehaene & Mehler, 1992). Alternative explanations of the size effect rely on this property of the numerical stimuli. One model proposes a connectionist network to account for various number processing phenomena (Verguts et al., 2005; Verguts & Van Opstal, 2014). The model introduces a hidden layer where not only the nodes representing a value are activated, but also the nodes for the neighboring values, generating the distance effect. Importantly, the activation of the neighboring nodes does not depend on the value of the numbers; in other words, the noise of the represented value has a fixed width. Unlike other ANS models, this representation is appropriate to explain the distance effect, but it cannot explain the size effect. To account for the size effect, uneven everyday frequencies of the numbers are introduced to the model (Verguts & Fias, 2004; Verguts et al., 2005). Overall, this connectionist model suggests that the size effect is a direct consequence of the uneven frequency of everyday numbers.

Another alternative account which suggests that the size effect is a frequency effect proposed that symbolic numbers could be processed by a discrete symbolic network, the Discrete Semantic System (DSS) (Krajcsi et al., 2016). In this model, the distance effect might be rooted in the associations between the numbers and the large-small concepts, where numbers with closer values have similar associations with the large-small concepts, resulting in a more difficult decision (Krajcsi et al., 2016). Importantly, in the DSS model, similar to the connectionist model described above, the size effect is a frequency effect (Krajcsi et al., 2016): Smaller numbers are more frequent in everyday life (Dehaene & Mehler, 1992), and because more frequent stimuli are easier to process, smaller numbers are also easier to process, resulting in the size effect. See more details about the DSS model and the comparison of the DSS and the connectionist model in Krajcsi et al. (2016).

It has recently been shown that the distance and size effects do not correlate in symbolic number comparison (Krajcsi, 2016). This result is in contrast with the ANS model: The model proposes that the two effects are simply two ways to measure the single ratio effect, therefore, the two effects should correlate. On the other hand, the result is in line with the connectionist and the DSS model, which models suppose that the two effects have different sources. In another study, it has been shown that for new artificial number digits, where the participants do not have former information about the frequencies of those digits, the size effect follows the frequency of the digits in that session (Krajcsi et al., 2016). To our knowledge this was the first study that empirically and directly investigated the role of the frequency of the symbols in the size effect. In that study it was found that when the frequency of the digits followed the frequency of the Indo-Arabic numbers in everyday life, the usual size effect was observed. However, when the frequency of the new symbols was uniform, the size effect did not appear, suggesting that the size effect is a frequency effect. Both of these studies support the connectionist and the DSS models, and provide new evidence that cannot be explained by the ANS model.

In the latter study, in which the size effect of the new symbols followed the frequency of those digits, new artificial symbols were utilized. Because it was supposed that while the former experience with the already well-known Indo-Arabic numbers could influence the perceived frequency, new symbols may lack this prior information. Thus, new symbols may start to accumulate frequency statistics from an empty initial state (Krajcsi et al., 2016). In line with this, most of the former studies that investigated the size effect in Indo-Arabic numbers, applied uniform frequency for the digits in the sessions, and to our knowledge none of them reported zero size effect. This means that the uniform distribution statistics of a session cannot, at least not entirely, overwrite the statistics of former life experience. However, there could be partial changes which may be observed as a gradual change throughout the session. To test this possibility, in a recent study, the size effect was investigated in three blocks within a session where participants compared Indo-Arabic number pairs and numbers were presented with uniform frequencies (Kojouharova & Krajcsi, 2018). No significant changes were found across the blocks of the session, even if the descriptive data showed a slight decrease in the size effect slope. This result is also interesting in contrast to the distance effect flexibility in Indo-Arabic numbers, which flexibility was also tested in the same experiment. It was found that the distance effect has already significantly changed in the first block of that single session. (In that study it was investigated whether the distance effect is based on the association between the digits and the "smaller"-"larger" categories. To test this property, the frequency of the association of the numbers with the "larger" category was manipulated, and the distance effect was already found to be specifically distorted according to the manipulated associations at the beginning of the session. See Kojouharova & Krajcsi (2018) for the exact nature of this association-based distortion of the distance effect.) Returning to the size effect, while the effect did not change significantly, there was a slight decrease in the descriptive data. This may be either simply a noise of random sampling, or it is possible that there has been some change in the size effect slope, though the effect size (i.e., the magnitude of the effect, and not the size effect, which is the effect related to the size of the stimuli) was not large enough to reach significance with that sample size. Unfortunately, we are not aware of any similar studies investigating the change of the size effect through time with uniform stimuli distribution. One way to increase the potentially small statistical power in the Kojouharova and Krajcsi (2018) study is to utilize a different design: we might increase the effect size by applying not only uniform frequency, but everyday-like frequency (frequent small numbers and rare large numbers) and reversed everyday-like frequency (rare small numbers and frequent large numbers) (see Figure 26).

The aim of the present study is to extend the findings of Krajcsi et al. (2016) and Kojouharova & Krajcsi (2018) and test the flexibility of the size effect in Indo-Arabic numbers: Whether the size effect of the Indo-Arabic numbers also depends on the frequency of the digits. We investigate whether everyday, uniform and reversed everyday number frequency of the session can change the size effect between

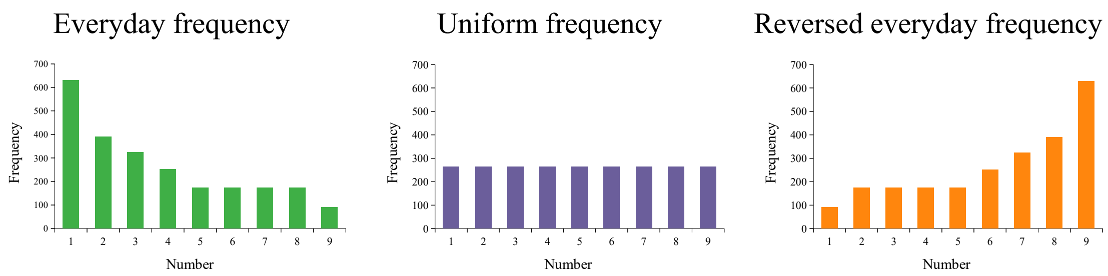


Figure 26. *Example frequencies with distribution similar to everyday life (Dehaene & Mehler, 1992), uniform distribution and a distribution of reversed everyday life. Frequencies of the values were generated based on the round (value^{-1} × 10) function.*

conditions and within session. Additionally, compared to the study by Kojouharova & Krajcsi (2018), the length of the session was increased from 300 trials/block to approximately 800 trials/block, to potentially increase signal-to-noise ratio. If the size effect can be changed even within a single session (as can be the distance effect, Kojouharova & Krajcsi, 2018)), then it would mean that frequency guides the size effect not only for new number symbols (Krajcsi et al., 2016), but also for Indo-Arabic numbers. This result would offer a more direct evidence for models highlighting the role of the frequencies in the number comparison size effects (Krajcsi et al., 2016; Verguts & Van Opstal, 2014).

Method

The participants compared Indo-Arabic numbers from 1 to 9 by selecting the larger member of a number pair. The numbers were presented in three different frequency conditions in a between-subjects experimental design.

Participants.

Forty-nine university students participated in the experiment for partial course credit. Because the effect of the frequency manipulation in Indo-Arabic numbers has not been investigated before, it is impossible to perform a power analysis. (Note that the effect the Kojouharova and Krajcsi (2018) study investigated was the change across the blocks within a session, and not the effect of frequency manipulation between conditions, therefore, the potentially small power of that study is not relevant in assessing the power of the present measurement.) We chose a sample size (approximately 15 participants per condition) that usually gives reliable distance and size effects in symbolic comparison tasks. One participant's data was not recorded properly, and was excluded from further analysis. Two further participants were excluded for having an error rate higher than mean + 2 standard deviations (higher than 16%). Thus, the data of 46 participants were analyzed (35 females, 21.02 years of mean age, 2.37 years *SD*): In the everyday frequency group 13 participants, 8 females, 20.31 years of mean age, 1.14 years *SD*; in the uniform frequency group 11 participants, 8 females, 21.64 years of mean age, 2.57 years *SD*; in the reversed everyday frequency group 22 participants, 19 females, 21.14 years of mean age, 2.68 years *SD*.[12] All but three participants were right-handed, and all had normal or corrected to normal vision. The present study was carried out in accordance with the recommendations of the Department of Cognitive Psychology ethics committee. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

Stimuli and procedure.

Pairs of Indo-Arabic numbers were presented to the participants in a number comparison task. In each trial, the participants decided which was the larger number by pressing the R and I keys of a keyboard (Figure 27). The two numbers remained visible until response. After the response a blank screen was visible for 300 ms, then the next trial started.

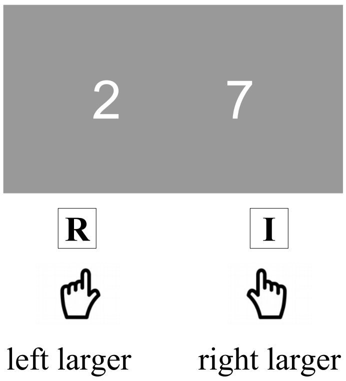


Figure 27. Number comparison task in the experiment. The participants selected of the larger number (number on the left or number on the right) of the presented pair by pressing a key.

12 The three frequency groups were planned to be equal in size, but because of an administrative error, the reversed everyday frequency group became larger than the other two groups. Still, the assignment of the participants was random. Importantly, the presented analyses handle the unequal size of the groups, and our replication study (see below) with equal group sizes also finds the same results as the present study, therefore, unequal group size cannot be an issue here.

Each condition consisted of three blocks for which the number pairs (trials) were generated as follows. In the everyday frequency condition the frequency of each number was calculated according to the $frequency_{value} = value^{-1} \times 10$ formula which yielded the following (rounded) frequencies (value:frequency): 1:10, 2:5, 3:4, 4:3, 5:2, 6:2, 7:2, 8:2, 9:1 (Figure 26) (see Dehaene & Mehler, 1992; Krajcsi et al., 2016). The 2-permutations of all numbers excluding ties were generated, resulting in 794 trials. Each pair was presented once in each of the three blocks for a total of 2382 trials. The same procedure was repeated for the reversed everyday frequency, but the frequencies were reversed before creating the number pairs, i.e., value:frequency: 1:1, 2:2, 3:2, 4:2, 5:2, 6:3, 7:4, 8:5, 9:10 (Figure 26). In the uniform frequency condition all possible pairs excluding ties were generated. Each pair was presented 11 times in a block, resulting in 792 trials per block and 2376 trials overall. The order of the trials within a block was random. There were no practice trials in the session. All participants were assigned randomly to the appropriate condition.

The presentation of the stimuli and recording of the responses were managed by the PsychoPy software (Peirce, 2007).

Data analysis.

In the present analysis we investigate the presence of the size effect throughout the experimental blocks (i.e., within a session) and the difference between the three frequency conditions in the whole stimulus space (i.e., between groups).

The size effect was investigated in the whole stimulus space. Expected patterns for the whole stimulus space if frequencies entirely influence the performance can be observed in Figure 28. Performance is depicted with columns and rows denoting the two numbers to be compared (all possible pairs, excluding ties), and cells containing performance. The size effect is analyzed in the whole stimulus space instead of the usual method (i.e., combining cells with the same size), because (a) it is more informative than simpler indexes of the effects, since any systematic deviation from the expected patterns of the effects could be observed, and (b) with the relatively large number of cells any systematic pattern could be convincing independent of the statistical hypotheses tests.

Figure 28 shows expected patterns in the three frequency conditions if frequency entirely influences performance. The figures display not only the size effect, but also the distance effect. The distance effect appears as an improvement in performance from the main diagonal towards the bottom-left and top-right corners[13], whereas the size effect appears as a change of performance along the main diagonal (top-left to bottom-right corner). In all cases, the size effect regressor was quantified as the sum of the two numbers. Size and distance effects are combined as the sum of the two effects.

Size effect is present

	1	2	3	4	5	6	7	8	9
1		-0.30	0.29	0.60	0.79	0.91	0.99	1.05	1.08
2	-0.30		-0.50	0.09	0.40	0.59	0.71	0.79	0.85
3	0.29	-0.50		-0.70	-0.11	0.20	0.39	0.51	0.59
4	0.60	0.09	-0.70		-0.90	-0.31	0.00	0.19	0.31
5	0.79	0.40	-0.11	-0.90		-1.10	-0.51	-0.20	-0.01
6	0.91	0.59	0.20	-0.31	-1.10		-1.30	-0.71	-0.40
7	0.99	0.71	0.39	0.00	-0.51	-1.30		-1.50	-0.91
8	1.05	0.79	0.51	0.19	-0.20	-0.71	-1.50		-1.70
9	1.08	0.85	0.59	0.31	-0.01	-0.40	-0.91	-1.70	

Size effect is absent

	1	2	3	4	5	6	7	8	9
1		0.00	0.69	1.10	1.39	1.61	1.79	1.95	2.08
2	0.00		0.00	0.69	1.10	1.39	1.61	1.79	1.95
3	0.69	0.00		0.00	0.69	1.10	1.39	1.61	1.79
4	1.10	0.69	0.00		0.00	0.69	1.10	1.39	1.61
5	1.39	1.10	0.69	0.00		0.00	0.69	1.10	1.39
6	1.61	1.39	1.10	0.69	0.00		0.00	0.69	1.10
7	1.79	1.61	1.39	1.10	0.69	0.00		0.00	0.69
8	1.95	1.79	1.61	1.39	1.10	0.69	0.00		0.00
9	2.08	1.95	1.79	1.61	1.39	1.10	0.69	0.00	

Reversed size effect is present

	1	2	3	4	5	6	7	8	9
1		0.30	1.09	1.60	1.99	2.31	2.59	2.85	3.08
2	0.30		0.50	1.29	1.80	2.19	2.51	2.79	3.05
3	1.09	0.50		0.70	1.49	2.00	2.39	2.71	2.99
4	1.60	1.29	0.70		0.90	1.69	2.20	2.59	2.91
5	1.99	1.80	1.49	0.90		1.10	1.89	2.40	2.79
6	2.31	2.19	2.00	1.69	1.10		1.30	2.09	2.60
7	2.59	2.51	2.39	2.20	1.89	1.30		1.50	2.29
8	2.85	2.79	2.71	2.59	2.40	2.09	1.50		1.70
9	3.08	3.05	2.99	2.91	2.79	2.60	2.29	1.70	

Figure 28. The expected distance and size effects in the three frequency conditions, if the frequency entirely influences performance. The distance effect regressor was calculated as log(large − small), where log is natural logarithm (see Footnote 14), and large and small are the large and small numbers of the pair. The size effect regressor was computed as the sum of the two values. The combined pattern was calculated according to the log(large − small) + a *(large + small) formula. Parameter* a *is set to −0.1 for the everyday frequency condition, to 0.1 for the reversed everyday frequency condition, and to 0 (i.e., it removes the size effect component from the function) for the uniform frequency condition. Darker shade indicates worse performance.*

In the present analysis, the size effect regressor was fitted to the error rates, reaction times and drift rates (see the next paragraph for details about the drift rate), then the slope of the size effect was tested between conditions and blocks.[14] First, for each condition, we calculated (a) the mean error rates, (b) the mean reaction times of the correct responses, and (c) the drift rates for each participant as well as for the whole

---

13 Here, logarithm of the distances instead of the linear distances was used, because (a) linear distance effect would cause negative performance values when the distance is very large, which performance would not make sense, and (b) previous data suggested that the distance effect can be described more appropriately with the logarithm function instead of the linear version (Krajcsi & Kojouharova, 2017; Krajcsi et al., 2016).

14 Note that while the distance effect is discussed in the Introduction and in the Discussion, because the ANS model supposes that both distance and size effects originate in the ratio effect, in the present test, which manipulates the frequencies of the digits, it is only the size effect that is relevant, and the distance effect is not investigated

stimulus space (i.e., for each presented number pair). In the case of reaction time, extreme values above 2000 ms were excluded which resulted in the removal of 0.57% of all trials (0.27% for the everyday frequency condition, 0.96% for the uniform frequency condition, and or 0.54% for the reversed everyday frequency condition).

Drift rate is an index from the increasingly popular diffusion model analysis, and is assumed to provide a more sensitive measure of performance (Ratcliff & McKoon, 2008; Smith & Ratcliff, 2004). In this model, evidence is accumulated gradually from perceptual and other systems until a sufficient amount of evidence becomes available for a decision to be made. For example, in a comparison task, the evidence is the information about which of the two numbers is larger. Drift rate represents the quality of information upon which the evidence is built, and while error rates and reaction times adequately capture performance on a task, drift rate is more directly related to the background mechanisms of performance. For example, in a comparison task, larger drift rate means a more solid information about which number is larger, which leads to faster and less erroneous responses. Drift rates can be recovered based on the observed error rate and reaction time parameters (Ratcliff & Tuerlinckx, 2002; Wagenmakers et al., 2007). While in a diffusion model other parameters than drift rate could also be recovered, only drift rate is investigated here, because it is the parameter that mostly reflects the difficulty of the relevant feature processing according to both the diffusion models (Smith & Ratcliff, 2004; Wagenmakers et al., 2007) and the ANS model (Dehaene, 2007). Here, we applied the EZ-diffusion model (Wagenmakers et al., 2007), a simplified version of the diffusion model which still allows for the recovery of drift rates from a relatively small number of trials in a cell. (In the present experiment, in the uniform frequency condition, a cell included 33 trials, and in the everyday and reversed everyday condition, a cell could include trials between 6 and 150, depending on the frequencies of the numbers.) For edge correction we used the half-trial solution (see the exact details about edge correction in Wagenmakers et al., 2007). The scaling parameter, within-trials variability of drift rate was set to 0.1 in line with the tradition of the diffusion analysis literature.[15]

15 Edge correction is used when the error rate is either 0%, 50% or 100%, and the formulas recovering the parameters would lead to an undefined operation result (see Wagenmakers et al., 2007 for the exact formulas). The within-trial variability can be set to any arbitrary values, because it only "scales" all other parameters without changing their relations, therefore, any values could be used. To make our results more comparable with other diffusion analysis studies, we use the usual 0.1 value as a scaling parameter.

An additional effect which may appear in a number comparison task is the end effect – reaction time is faster and error rate is lower for the cells containing the largest number of the set (Balakrishnan & Ashby, 1991; Piazza, Mechelli, Butterworth, & Price, 2002; Sathian et al., 1999; Scholz & Potts, 1974), in this case all cells with the number 9 in them (see Figure 29). This effect must be taken into account as it distorts the slopes of the size and distance effects in the concerned cells, i.e., the size effect slope would include not only the size effect, but also an unknown portion of the end effect. If the end effect is not linearly added to the distance and size effects, it cannot be disentangled with the help of multiple linear regression. Thus, if an end effect is present, an appropriate solution is the exclusion of the concerned cells from the statistical analysis. Therefore, in the present study, cells including the number 9 were omitted from the analysis.

Because the distance effect regressor and the size effect regressor (i.e., the predictors of our analysis) do not correlate at all (i.e., $r = 0.0$), and uncorrelated regressors do not influence each other's slopes, it was possible to measure the size effect with a simple linear regression, and there was no need for multiple linear regression.

Overall, size effect regressor was fitted to the error rate, reaction time, and drift rate data for each participant, and for the cells including numbers from 1 to 8 (i.e., cells not affected by the end effect). Then, the slopes of the size effects were tested against 0 with one-sample hypothesis test, and the difference between the conditions was also tested by comparing the slopes of the size effect across the three conditions. Finally, to examine whether the size effect decreased across the blocks, first, size effect slopes were calculated as above for all the three blocks. Then the change in the slope of the size effect as well as the difference between the conditions were compared in a 3×3 analysis of variance with blocks (Block 1, 2, and 3) and conditions (everyday frequency, uniform frequency, and reversed everyday frequency) as factors.

Results

The raw data of the measurement is available at https://osf.io/2hrms/. Figure 29 shows mean error rates, reaction times, and drift rates for the whole stimulus space for each condition. Figure 30 illustrates the size effect, already observable in Figure 29, in a more traditional way, which is more in line with the usual depiction of the distance and size effects in former works. (Note that the latter figure is only for illustration purposes, and the size effect is calculated on the whole stimulus space displayed on Figure 29, and as explained in the Methods section).

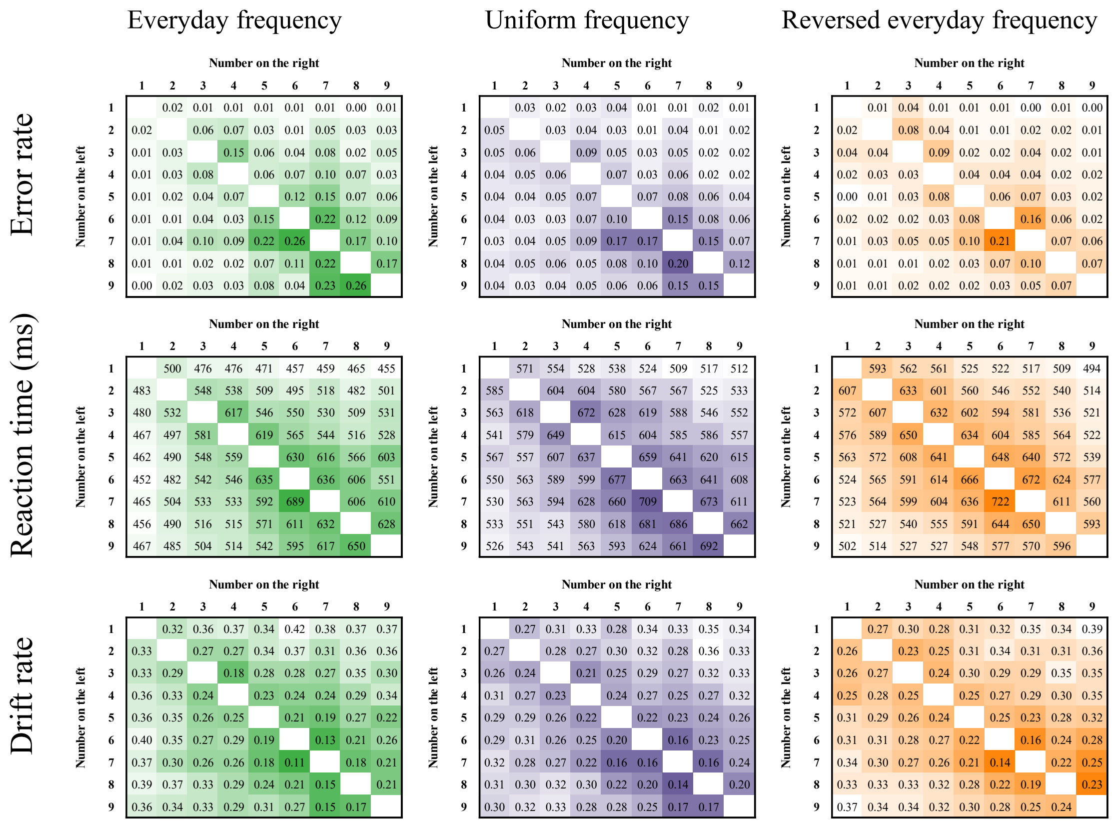


Figure 29. Average error rates, reaction times, and drift rates for the whole stimulus space for the everyday frequency, the uniform frequency, and the reversed everyday frequency conditions for all trials. Darker shade indicates worse performance.

Visual inspection suggests that there is a moderate end effect present in all conditions (see a description and present handling of the end effect in the Data analysis section), albeit to a different extent. Because the end effect was mostly present, the cells containing the number 9 were removed from further statistical analysis.

The size effect regressor was fitted with simple linear regression to the remaining cells, and the slope of the effect was tested against 0. The average slopes of the size effect regressor for all trials are presented on Figure 31. All slopes significantly deviated from 0 for error rates (everyday frequency: $Z = 3.180$, $p = 0.001$, tested with Wilcoxon signed-rank test, uniform frequency: $Z = 2.934$, $p = 0.003$, and reversed everyday frequency: $Z = 4.075$, $p < 0.001$), for reaction times (everyday frequency: $Z = 3.180$, $p = 0.001$, uniform frequency: $Z = 2.845$, $p = 0.004$, and reversed everyday frequency: $Z = 3.782$, $p < 0.001$), and for drift rates (everyday frequency: $Z = -3.180$, $p = 0.001$, uniform frequency: $Z = -2.485$, $p = 0.004$, and reversed everyday frequency: $Z = -3.425$, $p = 0.001$).

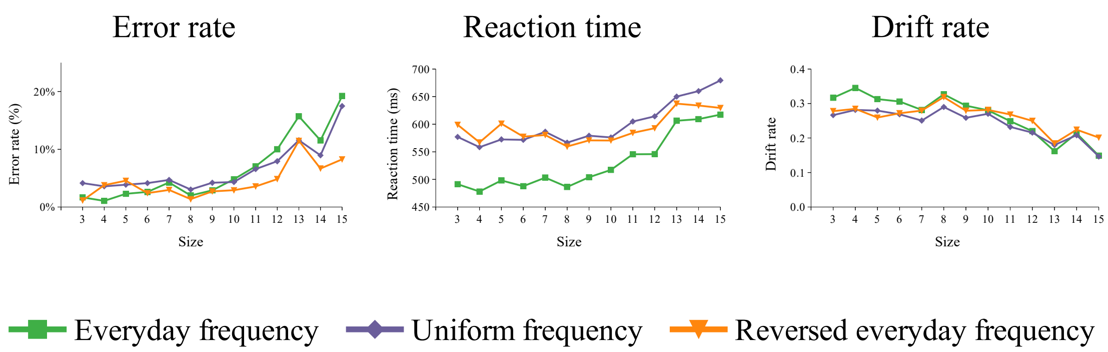


Figure 30. Size effect: performance as a function of size (sum of the two numbers to be compared) for error rates, reaction times, and drift rates. Note that number pairs containing the number 9 are excluded.

The differences of the size effect slopes between the frequency conditions were also significant: The size effect was largest for the everyday frequency, smaller for the uniform frequency, and smallest for the reversed everyday frequency condition (error rates: $\chi^2(2, N = 46) = 9.670$, $p = 0.008$, tested with Kruskal-Wallis test, reaction times: $\chi^2(2, N = 46) = 14.399$, $p = 0.001$, and drift rates: $\chi^2(2, N = 46) = 20.407$, $p < 0.001$). Dunn's post-hoc tests revealed that the slope of the size effect in the everyday frequency condition differed significantly from that in the reversed everyday frequency condition for error rates ($p = 0.006$), reaction times ($p = 0.001$), and drift rates ($p < 0.001$).

An additional analysis of variance was conducted for changes of size effect slopes with the progression of the task with factors for blocks (1st, 2nd and 3rd blocks) and frequency (everyday, uniform and reversed frequencies) conditions. The average slopes for each block and frequency condition can be seen on Figure 31. There was a main effect of frequency condition for error rates ($F(2, 43) = 8.882$, $p = 0.001$, $\eta_p^2 = 0.292$), reaction times ($F(2, 43) = 7.576$, $p = 0.002$, $\eta_p^2 = 0.261$), and drift rates ($F(2, 43) = 20.5507$, $p < 0.001$, $\eta_p^2 = 0.489$), repeating the result of the previous analysis on the role of frequency conditions. There was also a main effect of block only for error rates ($F(2, 86) = 9.4629$, $p < 0.001$, $\eta_p^2 = 0.180$) with less steep slopes in the first block than in the following two ($p = 0.014$ for the difference between Block 1 and Block 2 and $p < 0.001$ for the difference between Block 1 and Block 3, tested with the Tukey HSD post hoc test). The main effect of block was not significant either for the reaction times (F(2,

86) = 0.891, p = 0.414, η_p^2 = 0.020) or drift rates (F(2, 86) = 1.824, p = 0.168, η_p^2 = 0.041) and none of the interactions were significant (F(4, 86) = 1.154, p = 0.337, η_p^2 = 0.051, F(4, 86) = 0.714, p = 0.585, η_p^2 = 0.032, F(4, 86) = 1.905, p = 0.117, η_p^2 = 0.081 for error rates, reaction times and drift rates respectively).

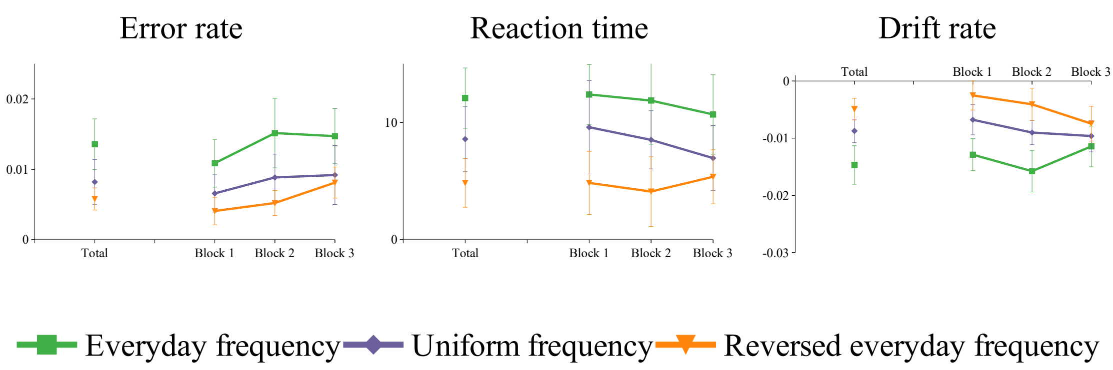


Figure 31. Average slopes of the size effect regressor for error rates, reaction times, and drift rates for all trials and for each block of the experimental session. Error bars show the 95% confidence intervals.

To summarize, the slope of the size effect was investigated in the whole stimulus space, excluding the cells containing the end effect. The frequency condition had an effect on the size effect, although the size effect did not disappear in the uniform condition or did not reverse (or even disappear) in the reversed everyday condition. Additionally, the effect of the frequency is observable even in the first block of the session, and the size of the effect did not change in the second and third blocks (except for the error rates where the effect size increased, but only between the first and the second blocks).

Replication study

In recent years, an increasing number of papers highlight the difficulty of replicating published results, revealing a distortion in published effect sizes and many times questioning whether the published phenomena are real (e.g., Open Science Collaboration, 2015). In line with these warnings, we rerun the previous experiment to see if the results are replicable.

Methods.

The stimuli, procedure and data analyses of the replication study was identical to the original experiment.

Thirty-two university students participated in the replication of the experiment for partial course credit. Two participants were excluded for having an error rate of approximately 50% (46.81% and 47.15%) which suggests chance performance. One additional participant was excluded for having an error rate higher than mean + 2 standard deviations (higher than 20.47%). Thus, the data of 29 participants were analyzed (18 females, 21.28 years of mean age, 1.98 years *SD*): In the everyday frequency group 9 participants, 3 females, 22.2 years of mean age, 1.69 years *SD*; in the uniform frequency group 10 participants, 6 females, 21.6 years of mean age, 1.63 years *SD*; in the reversed everyday frequency group 10 participants, 9 females, 20.1 years of mean age, 1.97 years *SD*. All but three participants were right-handed, and all had normal or corrected to normal vision.

Results.

Overall, the present results replicated the findings of the previous experiment.

The raw data of the measurement is available at https://osf.io/2hrms/.

The slopes of the size effect (calculated with single linear regression) significantly deviated from 0 for all frequency conditions for error rates, reaction times, and drift rates (all ps < 0.041) with the only exception of the reversed everyday frequency condition in the case of reaction times ($t(9) = 2.03$, $p = 0.073$).

The differences of the size effect slopes between the frequency conditions were significant for reaction times and drift rates; see Figure 32 (reaction times: $F(2, 26) = 3.608$, $p = 0.041$, $\eta_p^2 = 0.217$ and drift rates: $F(2, 26) = 7.81$, $p = 0.002$, $\eta_p^2 = 0.375$). Tukey HSD post-hoc tests revealed a significant difference between everyday frequency and reversed everyday frequency conditions in reaction time ($p = 0.032$) and in drift rate ($p = 0.002$). The size effect slopes followed the same tendency for error rates, but the difference was not significant with $F(2, 26) = 2.113$, $p = 0.141$, $\eta_p^2 = 0.140$.

The analysis of variance for changes of the size effect slopes between the blocks with factors for blocks (1st, 2nd, and 3rd blocks) and frequencies (everyday, uniform, and reversed everyday frequencies) conditions showed a tendency for the frequency main effect for reaction times ($F(2, 26) = 3.151$, $p = 0.059$, $\eta_p^2 = 0.195$) and a significant frequency main effect for drift rates ($F(2, 26) = 16.451$, $p < 0.001$, $\eta_p^2 = 0.559$), while the frequency main effect for error rates was not significant ($F(2, 26) = 2.113$, $p = 0.141$, $\eta_p^2 = 0.140$). Tukey HSD post-hoc tests showed that for reaction times the slope of the size effect in the everyday frequency condition was larger than that in the reversed everyday frequency condition (p = 0.048), whereas for drift rates the slope in the everyday frequency condition was larger than the slopes in the other two conditions (ps < 0.003). The main effect of the blocks ($F(2, 52) = 0.897$, $p = 0.414$, $\eta_p^2 = 0.033$, $F(2,52) = 0.696$, $p = 0.503$, $\eta_p^2 = 0.026$, $F(2, 52) = 0.413$, $p < 0.664$, $\eta_p^2 = 0.016$ for error rates, reaction times and drift rates respectively) and the interactions ($F(2, 52) = 0.329$, $p = 0.857$, $\eta_p^2 = 0.025$, $F(2, 52) = 0.962$, $p = 0.436$, $\eta_p^2 = 0.069$, $F(2, 52) = 1.812$, $p = 0.141$, $\eta_p^2 = 0.122$ for error rates, reaction times and drift rates respectively) were not significant.

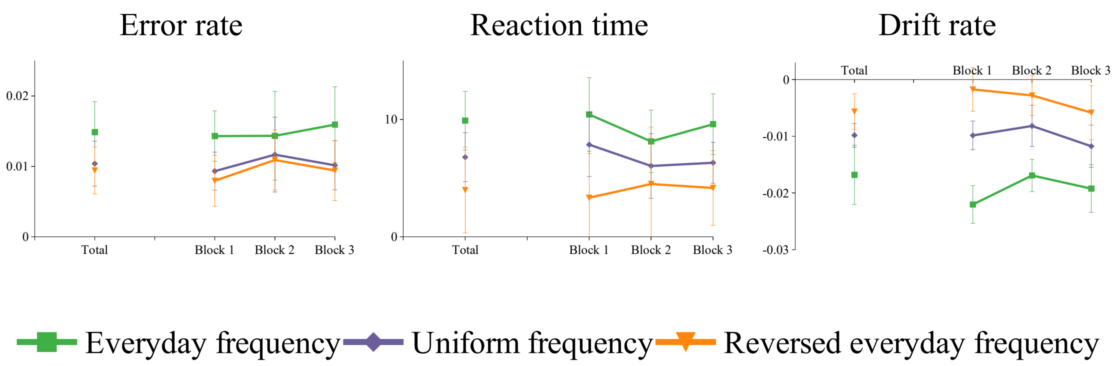


Figure 32. Average slopes of the size effect regressor for error rates, reaction times, and drift rates for all trials and for each block of the experimental session. Error bars show the 95% confidence intervals.

Discussion

The present study investigated the flexibility of the size effect in the Indo-Arabic number comparison task by testing whether the size effect can be changed within a single session. First, we found that the modified frequency of the stimuli of a single session can change the size effect. This was reflected in the significant size effect slope differences between the frequency conditions, suggesting that the size effect is flexible, and it can be modified even in a single session. However, we also found that the change is not complete, and with uniform frequency we still can see a significant positive slope, instead of a zero slope for the size effect, and even reversed everyday frequency is unable to reverse the size effect, or make it disappear (i.e., approaching zero slope). A final finding is that the effect of the frequencies can be observed during the first block of the session, however the size of the effect does not change throughout the second and third blocks. This latter aspect of the results may be considered as unusual: seemingly there is a very quick and efficient learning in the first block (as observed in the difference between the frequency conditions), but this learning is halted very quickly, and there is no change in the second and third blocks, even if the statistics of the session are not acquired entirely (reflected by the lack of interaction between the frequency condition and the block numbers). This seems to be an unusual result, because learning is often a gradual process, and typically an exponential learning function is observed, while here the learning is halted before the end of the first block. To summarize, we found that there is a fast, flexible component of the size effect, and there is also a stable component still kept throughout the session.

Contrasting the present results with similar previous studies, while with new symbols the size effect is entirely dependent on the frequency statistics of the session (Krajcsi et al., 2016), with Indo-Arabic numbers the size effect seems to be a mixture of the frequency statistics of the session and other effect(s) (as found in the present work). Also, while the frequency statistics of the session have an effect on the slope of the size effect as observed in the different slopes between conditions with different digit frequencies (present work), this change cannot be observed with the present power as a change within the session, because the change takes place at the very beginning of the session (Kojouharova & Krajcsi, 2018 and the present results).

Regarding the roots of the flexible component of the size effect, since the size effect changed with the experimental manipulation of the frequency, it is reasonable to assume that the flexible part of the size effect should be a frequency effect. What is more interesting and important is the root of the stable component of the size effect.

One possible source of the stable component of the size effect might be the ANS. The ANS model supposes a ratio-based effect that can also be observed as a distance or a size effect, and the stable part of the size effect can be that ratio effect. The presence of the distance effect is also in line with this account, which effect could also confirm the presence of the ratio effect. In this explanation, the stable and flexible parts of the size effect can come from different sources: the former is rooted in the ANS, and the latter is caused by the frequency of the stimuli. Although this explanation may account for the present results, it cannot explain other features of symbolic comparison: For example, the distance and size effects were found to be independent in symbolic comparison (Krajcsi, 2016); the distance effect can be present without the size effect when the frequency is uniform (Krajcsi et al., 2016); the distance effect follows the statistics of associations between numbers and small-large categories instead of the values of the numbers (Kojouharova & Krajcsi, 2018; Krajcsi & Kojouharova, 2017); unlike non-symbolic comparison performance, symbolic Indo-Arabic comparison performance cannot be described properly with psychophysical functions (Krajcsi, Lengyel, & Kojouharova, 2018). Overall, while the ANS model can explain the present results, it cannot explain the independence and flexibility of distance and size effects found in related studies.

Another possible explanation is that both stable and flexible parts of the size effect are frequency effects (i.e., the frequency of the stimuli are learned and reflected e.g., in processing time), and the seeming split of the size effect is only the consequence of a constrain in the learning process. It might seem unusual that instead of a gradual learning curve, frequency statistics of the session has an effect very quickly at the beginning of the session, and there are no further changes in the rest of the session, even if the statistics of the session is clearly not acquired entirely. This might seem unusual, because mostly it is supposed that the learning process is based on a gradual accumulation of environmental information, representationally updated by the delta rule. With this supposition, in the present data, either the statistics should show further changes in the second and third blocks, or the change in the first block should be smaller. However, recent findings reveal that learning might show different patterns not only by dynamically changing the learning rate (Behrens, Woolrich, Walton, & Rushworth, 2007), but for example, participants might consider perceived change in the environment, whether represented parameters should be updated (Arató, Khani, Rainer, & Fiser, submitted; Gallistel, Krishan, Liu, Miller, & Latham, 2014). Also, learning the statistics of an actual session would mean forgetting the statistics of all former experience, which is not always an ideal learning strategy, so occasionally some combination of the old and new information is required. Thus, it is not impossible that in the present data statistics of the session is integrated only partly into the former knowledge about digit statistics. Still, if this explanation is correct, it is not yet clear why the frequency statistics of the new session is utilized only partially. This is also an interesting question in light of the flexibility of the distance effect. As mentioned in the introduction, the distance effect might be the result of the association between the digits and the larger-smaller categories, and the statistics of the session is dominantly followed not only in new symbols (Krajcsi & Kojouharova, 2017) but already in the beginning of the session in Indo-Arabic numbers (Kojouharova & Krajcsi, 2018). In other words, unlike the size effect, distance effect follows dominantly the session statistics, ignoring former experience. Thus, if the frequency explanation is correct for the stable component of the size effect, it is not yet clear why the distance effect is more flexible than the size effect observed here.

There could be other explanations for the stable part of the size effect other than the ANS (Dehaene, 1992, 2007; Moyer & Landauer, 1967) or the models that propose the role of the frequency in the size effect, such as the connectionist model of symbolic number processing (Verguts et al., 2005; Verguts & Van Opstal, 2014) or the DSS model (Krajcsi, 2016; Krajcsi et al., 2016). However, we are not aware of any other model that could give a consistent explanation for the size effect.

The present work confirmed that symbolic comparison size effect is a frequency effect. This finding extends the idea that comparison distance and size effects are rooted in different mechanisms, not in a single ANS, as demonstrated by the independence of the symbolic distance and size effect slopes (Krajcsi, 2016), and by the fact that while the distance effect depends on the association of the digits and "small"-"large" categories (Kojouharova & Krajcsi, 2018; Krajcsi & Kojouharova, 2017), the size effect depends on the frequency of the digits (Krajcsi et al., 2016 and the present work). In a more applied line of research, the symbolic comparison task with distance effect slope is utilized routinely to predict everyday and school math performance (for a review, see e.g., Schneider et al., 2017), and it is usually interpreted that the sensitivity of the ANS is measured with this index. In the light of these recent findings, the role and validity of those indexes should be reconsidered. The comparison distance effect slope cannot be an ANS index, because in that case it also should measure the same construct as the size effect slope, which is seemingly not the case. Instead, symbolic distance and size effect slopes measure different constructs, and most probably none of them are related to the sensitivity of the ANS, but some properties of a symbolic system. Further research should find what exactly these properties are, and why they are good predictors of several math abilities and math performance indexes.

To summarize, the present experiment investigated the flexibility of the size effect in Indo-Arabic numbers, and it was found that the size effect is partly flexible and follows the frequency of the stimuli in the session. On the other hand, there is a stable part of the size effect, and importantly, while the session statistics is incorporated early in the session, no further change is observable in the rest of the session. While this restricted flexibility of the size effect can be explained by both the ANS model and the models proposing the role of the stimuli frequency in size effects, only the latter models can give a coherent view on the flexibility of distance and size effects in various symbolic number comparisons. Further research can determine whether the stable part of the size effect is also rooted in the frequency of the stimuli, and why size and distance effects have different flexibility.

Acknowledgments

We thank József Fiser for a discussion about the interpretation of the data. We thank Krisztián Kasos and Ákos Laczkó for their comments on an earlier version of the manuscript.

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

Discussion

Summary of the Results

Four studies investigated the source of the distance and size effects in symbolic numbers in the number comparison task, one of the most common paradigms in numerical cognition. As both effects are believed to be indicators of how the representation of numerosity works, this was also an indirect way of testing which of the proposed accounts provides a better explanation of the results from the manipulation utilized in those studies. On the one hand, the ANS, the Analogue Number System account, suggests a noisy, continuous, analogue representation that works according to Weber’s law, and both effects stem from the same source: The ratio of the values of the numbers that are compared. On the other hand, the DSS, the Discrete Semantic System account, supposes a semantic network in which the numbers are nodes, the distance effect is rooted in the strength of their connections or associations with other nodes, and the size effect depends on the everyday frequency of the numbers.

Thesis Study 1 began with a direct comparison the ANS and the DSS accounts. The models were quantified according to the available models in the literature in the former case, and possible quantification in the latter based on the known constraints. Then they were fitted with a linear regression to the performance in the number comparison task, which was measured in error rates, reaction times, and drift rate. For error rates and reaction times the results were inconclusive, with both models showing a similar fit where the advantage of one model over the other depended on the exact formulation of the quantification. For drift rates, the DSS seemed a better fit. Because of the relatively high noise and the uncertainties of the diffusion analysis method, this result can only be interpreted with caution. However, there were two conclusions that served as grounds for our further work: 1) a direct comparison of the ANS and the DSS is unlikely to be successful, so a different approach was required, and 2) there was not a clear preference for either of the models, which, at the very least, was an indication that the DSS could be as plausible an account for the distance and size effects as the ANS.

In an additional analysis, the presence of the distance and size effects was tested with both a multiple and a single linear regressions calculated for each participant and for the full stimulus space. The distance as a predictor was quantified as the absolute difference of the two numbers, and size was the sum of the two numbers. The slopes of the distance and size effect predictors from the linear regression fit were tested against 0, and the deviation was statistically significant for both effects and for both analyses. Along with confirming the trivial expectation that both effects should be present, this also meant that this type of analysis could be used in the rest of the studies. In Experiments 2 and 3 only multiple linear regression was used.

As both accounts offer putative sources for the distance and size effects, it was possible to turn to them as an indirect approach of comparing the two models. In Experiment 2 of Thesis Study 1 participants were taught new, artificial symbols for the numbers from 1 to 9. The reason for this choice was that the frequency of the new numbers can be manipulated when presented in a task unlike the everyday frequency of the Indo-Arabic digits. The new symbols were compared in the number comparison task. In one group the symbols were presented with uniform frequency, and in the other their frequency was biased towards the everyday frequency of Indo-Arabic digits (i. e., the new symbol for 1 was seen most frequently, the new symbol for 9 most rarely). Error rates and reaction times were analyzed. The slopes obtained from the multiple linear regression fit for the distance and size effects[16] showed that while the distance effect was present and similar in both groups, a size effect emerged only for the biased frequency condition. Moreover, the slopes of the size effect in the uniform frequency condition and in the biased frequency condition differed significantly from each other. Experiment 3 investigated a possible confound – the semantic congruence effect, according to which reaction times are faster for larger numbers when the instruction is “choose the larger”. This effect could have extinguished a possible size effect in the uniform frequency condition as it is in a direction opposite to that of the size effect. In Experiment 3 only the uniform frequency condition was run with a “choose the smaller number” instruction. The results of the multiple linear regression showed that the size effect was not present, and its slope in this condition was comparable to the slope in the uniform frequency condition in Experiment 2.

Crucially, the study utilized new symbols as numbers, so we had to ensure that the new symbols were treated as numbers. A priming task was included, in which the prime was always a new symbol and the target was always an Indo-Arabic digit. The presence of a priming distance effect, i. e., responses for the target are faster if the prime was semantically close (close in value) is considered to be an indicator of the same underlying representation or a representational overlap. While the results from this

16 For the end effect refer to the relevant analysis section in Thesis Study 1.

study alone lacked statistical power, a meta-analysis confirmed that a priming distance effect was present, i. e., the new symbols had the same meaning as the Indo-Arabic digits.

In Thesis Study 2 the source of the distance effect in symbolic number comparison was investigated. On the one hand, it could be rooted in the values of the numbers – a supposition that is in line with both the ANS and the DSS account. In the case of ANS, the distance effect is a direct result of the representation of the number. According to the DSS account, the distance effect could depend on the strength of the connections between the nodes, with stronger connections reflecting a closer meaning. On the other hand, it is also possible that the distance effect stems from the associations of the numbers with the "small-large" properties (also prominent in the delta-rule connectionist model (Verguts et al., 2005; Verguts & Van Opstal, 2014)). We quantified both predictions for the distance effect in the full stimulus space for a partial set of artificial numbers (1, 2, 3, 7, 8, 9), where the value-based model stated that the predictor for the performance for each number pair is the difference of their value, while the association-based model suggested that it is the difference between the values of their places in an ordered sequence (e.g., number 9 is in the 6th place in the sequence). Moreover, in both cases the linear difference and the logarithm of the difference were used as regressors, as there was an indication that the latter describes the distance effect somewhat better. Error rates, reaction times and drift rates were used in the analysis. A linear regression model fit at the group level indicated a better R^2 for the association-based model for reaction times and drift rates. The individual linear regression fit and the subsequent comparison of the individual R^2 provided further evidence for this conclusion. When the logarithm of the distance was used, the difference between the two models was larger and thus better defined. In addition to this study, a replication experiment was conducted. The R^2 result for error rates here clearly supported the value-based model at both group and individual level, while the results for the reaction times and drift rates were in line with the association-based model with the difference being significant at the group level. A meta-analysis including the original experiment, the replication experiment, and the experiment with Indo-Arabic digits from Thesis Study 3 was inconclusive in the case of error rates, but quite straightforward in the case of reaction times and drift rates: The association-based model was a better fit overall. Two additional results were that the lack of a size effect when artificial numbers are presented with uniform frequency replicated the result from Thesis Study 1, and that the logarithm of the difference (distance) between the numbers seems to be a better predictor than the linear difference.

Thesis Study 3 furthered the case for an association-based model for the distance effect by testing it against the value-based model in Indo-Arabic digits with the same method as in Thesis Study 2. We originally supposed that the associations formed with "small-large" properties in Indo-Arabic numbers would be stable, however, this presumption, i. e., the flexibility of the distance effect, had not been investigated before. Additionally, our results obtained with artificial numbers could still stem from a different source, thus an experiment in the Indo-Arabic notation could strengthen or weaken the evidence for an association-based account. This experiment also provided an opportunity to explore the flexibility of the size effect as numbers were presented with uniform frequency. Here, a decrease rather than full disappearance was expected, because the studies found in the literature that presented the stimuli with uniform frequency still produced the size effect. This was made possible by using a larger number of trials per stimulus pair (30), which were then divided into three blocks, so that a possible gradual change in both effects could be observed. Error rates, reaction times, and drift rates were analyzed in a multiple linear regression with the distance effect regressor being the logarithm of the absolute difference between the numbers (once for the value-based model and once for the association-based model) and the size effect regressor being the sum of the two numbers. The association-based model was a better fit in all cases, at the group and at the individual levels. Moreover, an analysis of variance for the possible change in the distance effect over the course of the experiment showed that the change happened fast, already in the first block. Regarding the size effect, it was present (its slope from the multiple linear regression significantly deviated from 0) for error rates, reaction times, and drift rates overall, with no change occurring between the blocks, although descriptively both an increase (error rates) and a decrease (reaction times) could be observed. The distance effect was reliable across blocks based on test-retest correlations between the first and the second block, and the second and the third block. The size effect was not reliable, but this may have been due to its low effect size. Thesis Study 3 further supported the use of the logarithm of the absolute distance as a predictor for the distance effect, and also suggested that the distance and size effects likely dissociate which argues against the idea of a common source.

Thesis Study 3 demonstrated that the distance effect is flexible in the Indo-Arabic notation. Although the study also examined possible changes in the size effect, it was not optimized for testing its possible flexibility. Building on the results for the flexibility of the distance effect in both artificial numbers and Indo-Arabic digits and our success in invoking a size effect in artificial numbers by introducing biased frequency, a study of the a possible flexibility of the size effect in the Indo-Arabic digits seemed justified (Thesis Study 4). Again, our presumption that the everyday frequency is well-established had not been investigated before. All Indo-Arabic digits (numbers 1 to 9) were used. In one condition they were presented with their everyday frequency, in the second with uniform frequency, and in the third with reversed everyday frequency. There were 30 trials per number pair which allowed for following the course of a possible change throughout the study by dividing the trials into three blocks. The slope of the size effect was significantly different from 0 for all three conditions for error rates, reaction times, and drift rates. It also significantly different between the three conditions for all three measurements. The slope, however, did not differ between the blocks except for error rates where it increased between the first and the second. That increase was present for all three conditions, thus it was likely driven by different factors. A replication experiment showed the same results for the slope of the size effect overall with the difference between conditions not being significant only for the error rates. The slope did not change between blocks. The replication study was underpowered, which may have contributed to a lack of statistical power for error rates, nevertheless, all results were in the same direction as in the original study. Overall, the results are in favor of either a two-component size effect, with one flexible and one stable component, or a partially flexible single component[17].

17 The latter explanation is explored, but not emphasized in Thesis Study 4.

Results and Aims

Here, I list once again the aims of the studies included in the thesis along with a statement based on the results:

1. examine whether a different model (DSS) is a better description for the data obtained in the symbolic number comparison task than the ANS (Thesis Study 1, Experiment 1);

   The results were inconclusive about which model is a better description, but provided evidence that the DSS is, at the very least, a plausible alternative model for explaining the distance and size effects in the number comparison task.

2. examine frequency as a possible source of the size effect by testing whether it can be induced by manipulating the frequency of presentation of the numbers when recently learned artificial numbers are used for which there is no prior experience (Thesis Study 1, Experiment 2 and 3);

   Manipulating the frequency of the numbers was sufficient to induce a size effect, which did not appear for the uniform frequency. Thus, frequency is the source of the size effect for new, artificial numbers.

3. examine the associations between the numbers and the “small-large” properties as a possible source of the distance effect by manipulating those associations in a new, artificial number sequence (Thesis Study 2);

   The association-based model was a better fit for the data obtained in this study, thus the associations between the numbers and the “small-large” properties are the source of the distance effect in the number comparison task for new, artificial numbers.

4. examine whether the associations between numbers and the “small-large” properties can be modified in Indo-Arabic numbers within a session of the comparison task, i.e., seek further confirmation for the distance effect being association-based (Thesis Study 3);

The association-based model was a better fit for the data in this study, thus the associations between the numbers and the "small-large" properties are the source of the distance effect in the number comparison task for Indo-Arabic numbers.

5. examine whether the size effect shows similar flexibility in Indo-Arabic numbers by manipulating the frequency of presentation of the numbers within a session, i.e. further confirmation for frequency being the source of the size effect (Thesis Study 3 and Thesis Study 4);

   The size effect was modified but not entirely removed by manipulating the frequency of the numbers in the number comparison task for Indo-Arabic numbers. Frequency nevertheless contributes to the size effect as a flexible component, and it is possible that everyday frequency could explain the stable component.

6. examine whether the distance and the size effects change independently of each other (all Thesis Studies);

   The distance and size effects changed independently as a result of the manipulation in the experiments, thus they are dissociated, i. e., they have a different source.

7. more generally, an aim present in all reported studies, contrast the two proposed models of numerical cognition, the ANS and the DSS, in symbolic numbers. Here, the sources of the numerical distance and size effects are examined for being consistent with either account, and conclusions about the two accounts will be drawn based on that, but any further investigation of the two models is beyond the scope of the thesis;

   The distance and size effects changed as a result of the experimental manipulation in a way predicted by the DSS and not predicted by the ANS. In this light, the DSS is the better account for symbolic numerical processing.

Theoretical Conclusions and Consequences

Source of the distance effect.

In both new symbols and the Indo-Arabic notation the distance effect was shown to be rooted in the associations between the numbers and the "small-large" properties (Thesis Study 2 and 3)[18]. The change occurred at the beginning of the session for the Indo-Arabic numbers, which points to the flexibility of the effect. The results are decisive in regards to reaction times and drift rates, confirmed also by the mini meta-analysis in Thesis Study 2. When measured with error rate, the results were equivocal, and it remains to be seen whether this was due to noise or any additional aspects of the distance effect. Nevertheless, the overall results[19] are a definite statement against a value-based explanation of the distance effect as suggested by the mainstream ANS model (Dehaene, 2007; Moyer & Landauer, 1967) or by the value-based version that can be formulated within the DSS model. The results are in line with the association-based explanation of the DSS model or by the delta-rule connectionist model of numerical effects (Verguts et al., 2005; Verguts & Van Opstal, 2014).

Source of the size effect.

The size effect was shown to be a consequence of the frequency manipulation of the numbers in new symbols (Thesis Study 1), and at least in part for the Indo-Arabic notation (Thesis Study 4). The effect did not take on entirely the session's statistics in the latter case, and could not be reversed or nullified even when the numbers were presented with reversed everyday frequency. It is possible that the size effect is a two-component effect, one of which is the frequency of the session and flexible, whereas the other is stable. Another possibility is that the size effect is rooted in the frequency with the numbers only partially acquiring the statistics of the session. In this respect the size effect is less flexible than the distance effect.

18 All results and interpretations here are under the understanding that both effects were investigated in the number comparison task; however, as performance in this task has extensively been used as proof of the automatic activation of the ANS in the case of symbolic numbers (e.g., Moyer & Landauer, 1967), this does not lessen the arguments against the ANS and the support for the DSS:

19 In my interpretation, the inconsistency regarding the error rate does not weaken the evidence against the ANS as according to the latter an inconsistency like this one should not be observable unless there is an additional mechanism that could distort the mapping between the symbolic and non-symbolic magnitudes proposed by the ANS. We argued against such a possibility in Krajcsi et al. (2018).

Here's the important point: No size effect was observed in Thesis Study 1 when the frequency of the numbers (new symbols) was uniform. It can be supposed that the putative stable component observed in Thesis Study 4 needs to be introduced into the system at some point during the time when numbers are learned. One possible source of this stable component of the size effect might be the ANS. That would mean that the size effect is a consequence of the ratio effect together with the distance effect. The presence of both effects in Thesis Study 4 is consistent with such a supposition. Thus, the size effect could have two sources: The ANS is the source for the stable component, and the frequency of the stimuli as the flexible component. However, 1) in Thesis Study 1 there is a distance effect when there is no size effect and the new symbols primed the Indo-Arabic numbers which supposes the same representation (or at least an overlap), 2) the distance effect does not reflect the value of the numbers as observed in Thesis Studies 2 and 3, and 3) the distance effect showed more flexibility by acquiring the statistics of a session in Thesis Study 3, while the size effect did not. If the stable component appears at a later stage through the involvement of the ANS in symbolic numerical cognition, that should mean either no distance effect before that, or there are two distance effects, and the ANS takes over or the two are combined. Since the same manipulation caused the same change in the distance effect in both new symbols and Indo-Arabic numbers (Thesis Studies 2 and 3), it is more likely that this is the same distance effect. Furthermore, there is no indication in our studies for a two-component distance effect. Some additional evidence is provided by the independence of the two effects (Krajcsi, 2016, also see below) as well as by another study of our laboratory (Krajcsi, Lengyel, & Kojouharova, 2018) demonstrating that unlike non-symbolic comparison performance, symbolic Indo-Arabic comparison performance cannot be described properly with psychophysical functions. To summarize, although the ANS is a plausible explanation for a two-component size effect, it cannot explain the additional findings from the Thesis Studies and other related studies.

The DSS offers an alternative explanation that can explain not only the partial flexibility, but can also accommodate the unusual result of a learning that does not occur gradually, i.e. gradual accumulation of environmental information, but very fast in the beginning with no changes later in the session. Here, the frequency of stimuli presentation is the cause of the size effect, and the observation that it is only partially flexible is a result of a constrain in the learning process. Recent works found that learning might show different patterns not only by dynamically changing the learning rate (Behrens et al., 2007), but for example, participants will likely deliberate whether the parameters of a perceived change in the environment should be updated (Arató et al., submitted; Gallistel et al., 2014). Moreover, a combination of old and new information seems the more reasonable strategy if it is uncertain whether and to what extent former experience needs to be updated. A partial integration of the statistics of the session is a sensible explanation for the present data. However, this raises further issues: Is there any specific reason as to why frequency (and possibly word frequency in general) only partially follows the session's statistics, and why is it that the distance effect (i.e., associations with a property) adopts it entirely from the beginning?

A third explanation for the stable component of the size effect besides the ANS (Dehaene, 2007; Dehaene, 1992; Moyer & Landauer, 1967), the DSS (Krajcsi, 2016; Krajcsi et al., 2016) or the delta-rule connectionist model (Verguts et al., 2005; Verguts & Van Opstal, 2014) is a definite possibility, but to our knowledge, no such explanation is available at present.

The distance and size effects are independent.

Further evidence that argues against a common (ratio-based) source of the distance and size effects is the independent change of the two effects. In Thesis Study 1 a size effect appeared only in the biased frequency condition, whereas the distance effect was present in both conditions (also in line with the delta-rule connectionist model). In Thesis Study 2, again, there was no size effect with the uniform frequency of stimuli presentation. In Thesis Study 3 the distance effect changed as a result of the manipulation, while the size effect was still observed, and in Thesis Study 4 the distance effect did not follow the change in frequency. An additional calculation in Thesis Study 3 shows a low correlation between the slopes of the two effects, thus further supporting an earlier correlational study by Krajcsi (2016) in which the slopes of the two effects were found not to correlate in symbolic numbers (Indo-Arabic digits), although they correlated highly for non-symbolic numbers.

Results regarding methodology.

Along with the theoretical implications for numerical cognition, there are also several methodological points that can be made based on our results. First, and as an entirely empirical result at this point, the distance effect is better described by the logarithm of the absolute difference of the values of the numbers to compare. This provides a more exact specification of the regressor, and is a step forward toward a possible quantification of the DSS model (or a part of the model).

The drift rate repeatedly proved to be the most sensitive index of performance in the comparison task, thus increasing support for the use of the diffusion model analysis as a more powerful method of analysis than the traditional error rate and reaction time.

By using the full stimulus space instead of collapsing the number pairs according to distance and size, we were able to use regressors for both effects simultaneously, recognize and handle a non-linear effect that influenced the results, and visualize our predictions and results in a more accessible manner. More importantly, with this data-driven approach it was possible to show that our results are observed systematic patterns in the data and not simply artifacts. In the cases where the traditional method of analysis is also depicted (Thesis Studies 3 and 4), the changes in the distance and the size effects can be clearly observed, however, because of the collapsing of cells into different categories depending on the prediction (more specifically in Thesis Study 3), it would have been more difficult to analyze the data appropriately.

The appearance of the priming distance effect in Thesis Study 1 between new symbols and Indo-Arabic digits, which interpreted according to the available information in the literature at present means that they shared a representation is a smaller but nonetheless important methodological point for future research that includes artificial numbers.

In some cases, we also included a replication study and applied mini meta-analyses in the studies. This is in line with current recommendations (e.g., Maner, 2014), and can contribute to a possible guideline in cases when the results of a study are inconsistent with the mainstream models and more evidence within one study contributes to making them more believable.

Last, one of the phenomena observable in the studies, but not discussed in detail, the end effect, may also have methodological and theoretical implications. The end effect means that decision is faster and with a lower error rate for number pairs containing the largest number of the set if the instruction is “which one is larger”, and there are very few studies that take this effect into account or use it from a theoretical point of view. For example, Pinhas and Tzelgov (2012) use the appearance of an end effect for the numbers 0 and 1, but not for the number 2 as an indication that the former two can serve as a starting point for the mental number line, while the latter cannot. Jou (2003) interprets it as those numbers being an anchoring point for making decisions. As a non-linear distortion of the data, it influences the analysis and the interpretation of the results. There are no guidelines as to how to handle the end effect. In the Thesis Studies, it was once included in a multiple linear regression and removed (Thesis Study 1), the cells containing the effect were removed (Thesis Studies 2 and 4), or it was ignored (Thesis Study 3). In all cases the decision was based on visual inspection of the effect and the extent of distortion. Future work may yield better methods for handling this effect.

ANS and DSS: Conclusions.

Based on the evidence obtained in the Thesis Studies regarding the distance and size effects, the following can be stated: 1) the source of the distance effect are the associations between the numbers and the "small-large" properties, 2) the source of the size effect is, at least in part, the frequency of the numbers, 3) the two effects are independent of each other, 4) the effects seem to be notation-dependent, i. e., the changes observable in both new symbols and Indo-Arabic numbers do not stem from the same representation (although they are caused by the same mechanisms: associations and frequency). The ANS and the DSS accounts make different predictions about the source of the two effects, thus, an indirect comparison is possible. According to the ANS the source of both effects is the ratio of the to-be-compared numbers, a consequence of their representational overlap. It also suggests a direct connection between a number and its value. The ANS model cannot explain the present results for symbolic numbers – any possible modification to align it with the data would mean changing a defining feature, the ratio-based performance, and so leading to an entirely new model.

The DSS, on the other hand, was able to present an alternative explanation that was comparable to the ANS model in the direct comparison as well as a better overall explanation for the changes in the distance and size effects caused by the experimental manipulation. With the same numerical effects being observed in both new symbols and Indo-Arabic numbers, a common mechanism for symbolic numerical processing is the parsimonious account. A further argument supporting this account would be that a system handling abstract symbolic operations such as the mental lexicon or a conceptual network is a reasonable proposal for symbolic numbers.

The DSS is an underspecified model. It relies on models describing higher-level cognitive (possibly linguistic) functions, so a quantitative description is not as readily available as the quantitative description (psychophysics formulas) of a low-level perceptual model such as the ANS. Whereas the ANS offers one explanation, the overlap of the representations of the numbers, for many of the properties of the numbers in different tasks, the DSS seemingly relies on several sources: connections between nodes, frequency, spreading activation. It is also more flexible, and may have several predictions about the outcome of an experiment. The latter can be seen in Thesis Study 2 where one of the predictions of the DSS is indistinguishable from the one of the ANS (value-based account), while the other contrasts it (association-based account). It is easy to focus on these shortcomings and miss the important points. First, the DSS is a comprehensible, cohesive account with precedents in the literature. Second, it is a more parsimonious account than the ANS for the processing of symbolic numbers: The organization of the system is responsible for the observed effects, i.e., the appearance of there being multiple sources is (in this sense only) illusory, and the effects can be observed for any processing of symbols in the relevant tasks. An example is the delta-rule connectionist model (Verguts et al., 2005; Verguts & Van Opstal, 2014), a possible implementation of the DSS, which accounts for belonging to a category (parity), the comparison distance effect, the priming distance effect, and the comparison size effect. Third, the DSS can explain all relevant symbolic numerical effects and phenomena at least as well as the ANS, and in some cases it provides a better explanation as seen, for example, in the present studies. Fourth, it can supply testable hypotheses not only against the ANS model, but also to contrast its own properties with one another, which in turn will result in a more precise description.

It is important to note that he DSS does not account for meaning. A different mechanism (presented fully by Krajcsi (2014)) suggests that initially number understanding is acquired via an object-based conceptual representation of natural numbers, stored and processed as exact values in an abstract layer. Further properties of the numbers can later be anchored in other domains (e.g., incorporating the meaning of zero as “nothing” (Krajcsi, Kojouharova, & Lengyel, in preparation), moving along a line for negative numbers (Krajcsi, 2014)) to extend their meaning. In this account the ANS is likely one of the systems in which number knowledge is grounded.

To sum up, the results about the distance and size effects in the Thesis Studies are already a step forward for a more precise specification of the DSS. Additionally, the results call for a re-evaluation of the interpretations from earlier studies based on the comparison of symbolic numbers. Further research is still needed to determine the feasibility of the DSS and to reveal and define its properties.

Other research supporting a distinct system for symbolic numerical cognition.

Despite the recent increased interest in the ANS and its relation to symbolic numerical processing, there are hardly any studies that challenge the nature of the distance and size effects in the number comparison task with direct manipulation as in the Thesis Studies (especially the size effect is rarely studied on its own). The results of Krajcsi (2016) revealed that the slopes of distance and size effects correlate highly in non-symbolic comparison, but do not correlate in symbolic comparison. The latter finding was also replicated in Thesis Study 3. Based on earlier studies Chesney (2018) modeled the distance effect for non-symbolic numbers and examined its relationship to the Weber fractions typically found in human subjects. The relationship was not linear, but rather J-shaped, and the distance effect correlated best with the Weber fraction at around 0.2-0.3 which is often not the Weber fraction range reported in studies. Thus, the distance effect is not even an optimal measure for ANS acuity, especially in special groups.

Nevertheless, more and more converging evidence is emerging in different tasks and experimental settings that supports the existence of a mechanism devoted to symbolic numerical processing. Some of the studies were discussed as a starting point for the Thesis Studies, grouped roughly as inconsistency with the ANS accounts in tasks when results for non-symbolic and symbolic numbers were compared, neurological evidence, predictiveness of non-symbolic and symbolic number processing for math achievement, and numerical-like effects in non-numerical tasks. A few additional studies are discussed as well below.

First, earlier studies found that the physical similarity of symbolic numbers contributed to the distance effect in a same/different task (Cohen, 2009; Ganor-Stern & Tzelgov, 2008). To avoid this problem Sasanguie et al. (2017) developed an audio-visual paradigm in which participants matched a number word or a tone sequence to a set of dots or a digit, and measured the ratio effect. In their first experiment there was no ratio effect in the pure symbolic (number word-digit) task, but it was observable in all tasks that included non-symbolic stimuli. In their second experiment, to rule out an asemantic route (i. e., meaning was not assessed in the pure symbolic condition), a comparison task was utilized. There was a reversed ratio effect only for reaction time in the number word-digit condition which contradicts the ANS account for an amodal common representation, although this result may have been confounded by an end effect. Also, interestingly, there was a correlation for the number word-digit and a sound-letter condition, a non-numerical condition. The audio-visual paradigm was applied in another study (Marinova, Sasanguie, & Reynvoet, 2018) to resolve a contradiction regarding the presence of a switching cost between notations that was observed in an earlier study (Lyons, Ansari, & Beilock, 2012), but not in the first two experiments of this study. Seemingly, the problem for this inconsistency was the faster reaction time for digits (in a dots-to-digit condition) than dots (in a digit-to-dots condition). The audio-visual paradigm kept the same "respond-to" stimulus independent of whether the trial was mixed or pure, and a switching cost, interpreted as an indication of two distinct systems for symbolic and non-symbolic numerical processing, was obtained.

There are also numerous examples of interference of discrete properties that are not easily explained by the ANS account. The numerical congruence effect (Henik & Tzelgov, 1982) may be attributed to an interference between (the continuous) numerosity and physical size as well as to the discrete "small-large" properties with which both numerosity and physical size can be described, i.e., the "numerical small-large" and the "physical small-large" properties interfere. The SNARC effect (responding to small numbers faster with the left hand and to large numbers with right hand) does not necessarily imply a space-numerosity interaction – it is possible that the "left-right" and the "small-large" properties interfere with each other. The SNARC effect was shown to appear when a spoken "yes-no" response, which should not have a spatial property, was required (Landy et al., 2008). In another study (Leth-Steensen et al., 2011) numerosity was associated with the "cold-warm" properties. Krajcsi et al. (2018) investigated the interference of the discrete properties parity ("odd-even"), side ("left-right") and numerosity ("small-large"). They showed that contradictions in earlier studies in regards to the interference of these properties may have been due to measurement – whether an effect is homogeneous like SNARC (the interference between the properties is similar for all participants) or heterogeneous like the other two (the interference was in different directions depending on the participant, but was consistent for the participant). Only discrete models are consistent with these results.

More precise methods and diverse designs seem to improve the results in neuroimaging studies. Event-related potentials (ERPs) remain a controversial approach mainly because of the difficulty of controlling for visual properties in both non-symbolic and symbolic stimuli. A recent study (van Hoogmoed & Kroesbergen, 2018) used a match-to-sample same/different task (match a target to a prime), where the prime and the target could be either a symbolic number or a set of dots. There was no ratio effect for the pure symbolic prime-target pairs for accuracy. No effect of ratio was found in the ERP waves for the difference between the prime and target in either of the conditions. Thus, it was not possible to compare directly the ratio effect. Still, the authors found differences in the processing of non-symbolic stimuli in the mixed conditions compared to the respective stimuli in the pure non-symbolic condition, and reasoned that because the differences were in amplitude, they were the result of different cognitive processes. Their conclusion was that symbolic numbers do not map onto the ANS. However, the study used double-digit numbers as stimuli which may have had a confounding effect on the results, i.e., additional non-numerical processes may have played a role.

fMRI (functional magnetic resonance imaging) studies seem to yield more promising results (for a review see Matejko & Ansari, 2018). Lyons and colleagues (2015) investigated the patterns of voxelwise correlations between pairs of numbers which should mirror the amount of overlap in their tuning curves. A match-to-sample task was used with the 1 to 9 numerosities. The expected overlap was found for non-symbolic numbers, but not for symbolic numbers whose pattern was more consistent with discrete representations. There was across-format correlation that did not show evidence of shared representation. However, the dot-array stimuli included the subitizing range (1 to 4) which may have affected the results. Two already mentioned studies (Bulthé et al., 2014; Bulthé et al., 2015) showed that symbolic numbers 1) likely recruit the entire cortex, 2) do not show a neural distance effect whereas dot arrays do, 3) do not have overlapping representations with non-symbolic numerosities in the selected regions of interest, 4) are recognized as an object (i. e., one digit is one object) in the IPS (intraparietal sulcus). The same authors (Bulthé, De Smedt, & Op de Beeck, 2018) recently published a study in which three groups were compared in a symbolic and non-symbolic comparison task. The groups differed in levels of arithmetic skills and experience: high, average, and low (developmental dyscalculia). Interestingly, there was a negative correlation between the MVPA (multi-voxel pattern analysis) generalization between symbolic and non-symbolic numbers and arithmetic performance in the parietal cortex. This result could be interpreted as a gradual separation (estrangement) between ANS and symbolic numerical processing, but could also entail a change in the representations of the symbolic numbers only.

Finally, a recent and rather novel study by Amalric and Dehaene (2017) targeted arithmetic skills in experts (mathematicians) and non-experts. They reported the results only for the expert group. In this group the mathematical and non-mathematical semantics separated, and mathematical semantics activated the IPS in a way similar to basic arithmetic and number recognition. This can be interpreted as higher arithmetic relying on the ANS or a similar mechanism. However, as the authors also recognize, the separation does not mean isolation from the language system, and they acknowledge the possibility for an organization similar to that of language. Furthermore, the IPS has been shown to be active for ordered sequences (Marshuetz et al., 2006; Matejko, Hutchison, & Ansari, 2018), and performance on ordinality tasks in turn is a good predictor of math achievement (e..g, Sasanguie, Lyons, De Smedt, & Reynvoet, 2017).

An interest in the role of ordinality in numerical cognition has recently re-surged (for a detailed review see Lyons, Vogel, & Ansari, 2016). The typical paradigm for measuring the role of ordinality is presenting pairs or triplets of stimuli (numbers, letters) either in ascending, descending, or mixed order, and participants make decisions about their order (ascending/descending or correct/incorrect). A reversed distance effect is usually observed when the digits are in ascending or descending order, i. e., performance is worse when the distance between the stimuli is large (Lyons & Beilock, 2013; Matejko et al., 2018; Sasanguie et al., 2017; Turconi, Campbell, & Seron, 2006 but see Experiment 1 in Sasanguie et al., 2017 for an exception). The usual numerical distance effect is obtained for mixed order trials (Lyons & Beilock, 2013; Matejko et al., 2018), but seems rarely to be the target of investigation. Behaviorally, a reversed distance effect was obtained in an ordinality task only for symbolic, but not for non-symbolic numbers (Lyons & Beilock, 2013). Goffin and Ansari (2016) showed that it did not correlate with the distance effect from the number comparison task, and was a unique predictor of math achievement along with the distance effect. A recent study by Sasanguie and colleagues (2017) found that when the ordering task used pairs, then performance on the comparison task predicted arithmetic performance, but was partially mediated by performance on the ordering task and fully mediated when performance on a letter ordering task was included in the model. In the same study when the ordering task included triplets, then it fully mediated between comparison and arithmetic, while letter ordering did not. From neurological point of view, a dissociation between cardinality and ordinality has been found (Lyons & Beilock, 2013; Matejko et al., 2018), although the results are somewhat inconsistent about whether there is a difference between numerical and non-numerical order processing and where it is. The parietal (and more specifically the IPS), prefrontal and even premotor regions have been implicated for both (Franklin & Jonides, 2009; Fulbright, Manson, Skudlarski, Lacadie, & Gore, 2003; Ischebeck et al., 2008; Lyons & Beilock, 2013; Marshuetz et al., 2006), although more sensitive analysis are likely necessary to investigate whether there really is an overlap between the brain networks (Zorzi, Di Bono, & Fias, 2011). Overall, ordinality seems to be distinct from cardinality, not limited to numbers, and is a major contributor to performance in symbolic numerical tasks.

Moving on to the adequacy of performance on the symbolic and non-symbolic number comparison task as a predictor of math achievement as investigated in developmental studies, we have already discussed that symbolic comparison distance effect seems to be a better predictor for math achievement in children than non-symbolic comparison performance (e.g., (Holloway & Ansari, 2009; Sasanguie et al., 2014, 2013). Additionally, there is a developmental trajectory:1) the symbolic distance effect correlates stronger for younger children (Holloway & Ansari, 2009) which underlines its importance in learning basic numeric skills; 2) children show greater variability for both non-symbolic and symbolic comparison in younger age which is likely due to a general noisier cognitive processes, and which may blur the difference between the performance in the two comparison tasks (Lyons, Nuerk, & Ansari, 2015); 3) in Grade 1 (around the age of six) the development of symbolic skills outpaces that of non-symbolic skills, and the refinement of non-symbolic skills seems to be guided by the symbolic skills (Matejko & Ansari, 2016), and 4) the latter reverse connection can be also observed by in children with dyscalculia – symbolic numerical impairment is displayed throughout their development, while ANS impairment appears only at 10 years of age and later (Noël & Rousselle, 2011). A meta-analysis study by Schneider et al. (2017) non-symbolic comparison correlates much less with math achievement, but a correlation with the number comparison task was found repeatedly, and another very recent study (Schwenk et al., 2017) showed that when comparison performance was measured with reaction time, children with mathematical difficulties were worse in both non-symbolic and symbolic comparison compared to typically achieving children, but to a greater extent in symbolic comparison. Factors such as age, diagnostic cutoff, number range did not affect the difference. These results suggest that if there is mapping between the symbolic numbers and ANS, it happens later in life, correlates with education, and is guided by the symbolic numerical knowledge rather than the other way around.

The DSS account, in contrast to the ANS model, can easily accommodate these results. As a mechanism dedicated to the processing of symbolic numbers, it is reasonable to assume that proficiency with tasks targeting basic skills such as the comparison task will predict better performance in more complex skills that are built on this foundation. The strength of the connections between the nodes representing numbers and properties can explain the interference effects such as the SNARC effect or the size-congruity effect, and their flexibility can account for the fast changes observed in the distance and size effects in the presented studies. The strong evidence about a semantics network and non-numerical properties in the IPS brain region also supports the DSS account, although this supposition still has to be treated with great caution. Obtaining different results in different tasks is also in line with the DSS – different parts of the network could be used depending on the task. The lack of a distance effect in the same/different tasks when an audio-visual paradigm is used is a curious deviation from the prediction of DSS if we suppose that the distance effect in these tasks is a result of a spreading activation between the nodes. One possible explanation is that number words and digits are not represented by the same nodes, thus there is no interference or facilitation. However, when visually presented, number words and Indo-Arabic digits show a distance effect in a same/different task (Dehaene & Akhavein, 1995), and prime each other in a priming task (Dehaene et al., 1998), i. e., there is a shared representation. Another unresolved question is that of the role of ordinality. The studies discussed here suppose that it is rooted in the connections between the symbols (nodes), which is possible in the DSS. However, if ordinality mediates between number comparison and arithmetic (Sasanguie et al., 2017), it might stem from the associations of the digits with the “small-large” properties which also gives the impression of an order-like quality.

Future Research and Practical Implications

From the point of view of the numerical cognition research, an important implication is that results obtained and interpreted on the basis of the distance and size effects need to be re-evaluated. Another line for future research is identifying and describing the properties of the DSS, which also means a more precise quantification of the account. Furthermore, the accumulated knowledge of symbolic numerical processing could be applied to language, e.g., acquiring of meaning, forming associations with properties, learning the statistics of the environment. Also, if Amalric and Dehaene’s (2017) data about a specialized mathematical semantic network is further corroborated, this could inform the forming of specialized language networks as well. On an even more general note, the present results can be extended to any field that seeks domain-specific evolutionary mechanisms for higher-level knowledge to be grounded in, more specifically, that the interpretation of such results in such way should always be treated with caution (see Núñez, 2017).

Application in practice could target mathematical education and intervention for children and adults with mathematical disabilities by developing new methods for teaching and testing, or providing a support system, tools that can help to process numerical information.

Conclusions

The studies presented in the thesis systematically investigated two phenomena stated to be an indicator of the underlying representation of numerosity and thus used to draw conclusions about its nature: the distance effect and the size effect in the number comparison task. Unlike other studies attempting to disentangle the non-symbolic and symbolic numerical cognition, we targeted directly the source of the two effects, or in other words, the numerosity representation. Our results were incompatible with the mainstream ANS account, which supposes an innate, analogue, noisy, continuous representation with evolutionary roots and in which the distance and size effects have the same source, the overlap of the number representations. The results are compatible with a representation in a semantic network such as the DSS as proposed in Krajcsi (2016) and Krajcsi et al. (2016) where the two effects are independent of each other. The distance effect is rooted in the associations of the numbers with the “small-large” properties, and the size effect is caused by the frequency of the numbers. By establishing the sources of the two effects it was possible to indirectly compare the two accounts. Additionally, the studies contributed to the quantitative description of the DSS, and provided methodological suggestions for future research. The studies thus become a part of the ever-growing body of research that distinguishes between symbolic and non-symbolic numerical cognition.

Table of Figures

Index of Tables

References

Amalric, M., & Dehaene, S. (2017). Cortical circuits for mathematical knowledge: evidence for a major subdivision within the brain's semantic networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1740), 20160515. https://doi.org/10.1098/rstb.2016.0515

Arató, J., Khani, A., Rainer, G., & Fiser, J. (submitted). *When more is less: Complex long-term serial effects in perceptual decision making*.

Balakrishnan, J. D., & Ashby, F. G. (1991). Is subitizing a unique numerical ability? *Perception & Psychophysics*, *50*(6), 555–564. https://doi.org/10.3758/BF03207540

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Beran, M. J. (2008). The Evolutionary and Developmental Foundations of Mathematics. *PLoS Biology*, *6*(2), e19. https://doi.org/10.1371/journal.pbio.0060019

Beran, M. J. (2012). Quantity judgments of auditory and visual stimuli by chimpanzees (Pan troglodytes). *Journal of Experimental Psychology: Animal Behavior Processes*, *38*(1), 23–29. https://doi.org/10.1037/a0024965

Beran, M. J., Johnson-Pynn, J. S., & Ready, C. (2011). Comparing children's *Homo sapiens* and chimpanzees' *Pan troglodytes* quantity judgments of sequentially presented sets of items. *Current Zoology*, *57*(4), 419–428. https://doi.org/10.1093/czoolo/57.4.419

Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1525), 1831–1840. https://doi.org/10.1098/rstb.2009.0028

Bulthé, J., De Smedt, B., & Op de Beeck, H. P. (2014). Format-dependent representations of symbolic and non-symbolic numbers in the human cortex as revealed by multi-voxel pattern analyses. *NeuroImage*, *87*, 311–322. https://doi.org/10.1016/j.neuroimage.2013.10.049

Bulthé, J., De Smedt, B., & Op de Beeck, H. P. (2015). Visual Number Beats Abstract Numerical Magnitude: Format-dependent Representation of Arabic Digits and Dot Patterns in Human Parietal Cortex. *Journal of Cognitive Neuroscience*, *27*(7), 1376–1387. https://doi.org/10.1162/jocn_a_00787

Bulthé, J., De Smedt, B., & Op de Beeck, H. P. (2018). Arithmetic skills correlate negatively with the overlap of symbolic and non-symbolic number representations in the brain. *Cortex*, *101*, 306–308. https://doi.org/10.1016/j.cortex.2018.01.008

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

Changizi, M. A., & Shimojo, S. (2005). Character complexity and redundancy in writing systems over human history. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1560), 267–275. https://doi.org/10.1098/rspb.2004.2942

Chesney, D. (2018). Numerical distance effect size is a poor metric of approximate number system acuity. *Attention, Perception, & Psychophysics*, *80*(5), 1057–1063. https://doi.org/10.3758/s13414-018-1515-x

Cohen, D. J. (2009). Integers do not automatically activate their quantity representation. *Psychonomic Bulletin & Review*, *16*(2), 332–336. https://doi.org/10.3758/PBR.16.2.332

Cohen Kadosh, R., Soskic, S., Iuculano, T., Kanai, R., & Walsh, V. (2010). Modulating Neuronal Activity Produces Specific and Long-Lasting Changes in Numerical Competence. *Current Biology*, *20*(22), 2016–2020. https://doi.org/10.1016/j.cub.2010.10.007

Crossman, E. R. F. W. (1955). The measurement of discriminability. *Quarterly Journal of Experimental Psychology*, *7*(4), 176–195.

https://doi.org/10.1080/17470215508416692

Cumming, G. (2012). *Understanding the new statistics: effect sizes, confidence intervals, and meta-analysis*. New York: Routledge, Taylor & Francis Group.

Dakin, S. C., Tibber, M. S., Greenwood, J. A., Kingdom, F. A. A., & Morgan, M. J. (2011). A common visual metric for approximate number and density. *Proceedings of the National Academy of Sciences*, *108*(49), 19552–19557. https://doi.org/10.1073/pnas.1113195108

Damarla, S. R., & Just, M. A. (2013). Decoding the representation of numerical values from brain activation patterns: Decoding the Representation of Numerical Values from Brain Activation Patterns. *Human Brain Mapping*, *34*(10), 2624–2634. https://doi.org/10.1002/hbm.22087

Dehaene. (2007). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In P. Haggard, Y. Rossetti, & M. Kawato (Eds.), *Sensorimotor foundations of higher cognition*: *Vol. XXII* (pp. 527– 574). Harvard University Press.

Dehaene, S., & Akhavein, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *21*(2), 314–326.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or Linear? Distinct Intuitions of the Number Scale in Western and Amazonian Indigene Cultures. *Science*, *320*(5880), 1217–1220. https://doi.org/10.1126/science.1156540

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1–2), 1–42. https://doi.org/10.1016/0010-0277(92)90049-N

Dehaene, Stanislas. (2011). *The number sense: how the mind creates mathematics* (Rev. and updated ed). New York: Oxford University Press.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371–396. https://doi.org/10.1037/0096-3445.122.3.371

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, *21*(8), 355–361. https://doi.org/10.1016/S0166-2236(98)01263-6

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, *43*(1), 1–29. https://doi.org/10.1016/0010-0277(92)90030-L

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., … Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, *395*(6702), 597–600. https://doi.org/10.1038/26967

Eger, E., Sterzer, P., Russ, M. O., Giraud, A.-L., & Kleinschmidt, A. (2003). A Supramodal Number Representation in Human Intraparietal Cortex. *Neuron*, *37*(4), 719–726. https://doi.org/10.1016/S0896-6273(03)00036-9

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fechner, G. T. (1860/1912). Elements of Psychophysics: Sections VII and XVI. In B. Rand (Ed.), & H. S. Langfeld (Trans.), *The Classical Psychologists* (pp. 562–572). Retrieved from http://www.lr-web.dk/Lru/microsites/hvadermatematik/hem2download/kap4_QR10_ekstra_fechners_originale_vaerk.pdf

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

Fischer, M. H., Mills, R. A., & Shaki, S. (2010). How to cook a SNARC: Number placement in text rapidly changes spatial–numerical associations. *Brain and Cognition*, *72*(3), 333–336. https://doi.org/10.1016/j.bandc.2009.10.010

Franklin, M. S., & Jonides, J. (2009). Order and Magnitude Share a Common Representation in Parietal Cortex. *Journal of Cognitive Neuroscience*, *21*(11), 2114–2120. https://doi.org/10.1162/jocn.2008.21181

Fulbright, R. K., Manson, S. C., Skudlarski, P., Lacadie, C. M., & Gore, J. C. (2003).

Quantity determination and the distance effect with letters, numbers, and shapes: a functional MR imaging study of number processing. *AJNR. American Journal of Neuroradiology*, *24*(2), 193–200.

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

Gallistel, C. R., Krishan, M., Liu, Y., Miller, R., & Latham, P. E. (2014). The perception of probability. *Psychological Review*, *121*(1), 96–123. https://doi.org/10.1037/a0035232

Galton, F. (1880). Visualised Numerals. *Nature*, *21*(533), 252–256. https://doi.org/10.1038/021252a0

Ganor-Stern, D., & Tzelgov, J. (2008). Across-notation automatic numerical processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(2), 430–437. https://doi.org/10.1037/0278-7393.34.2.430

Gebuis, T., & Reynvoet, B. (2012a). Continuous visual properties explain neural responses to nonsymbolic number: The neural correlates of nonsymbolic number. *Psychophysiology*, *49*(11), 1649–1659. https://doi.org/10.1111/j.1469-8986.2012.01461.x

Gebuis, T., & Reynvoet, B. (2012b). The Role of Visual Information in Numerosity Estimation. *PLoS ONE*, *7*(5), e37426. https://doi.org/10.1371/journal.pone.0037426

Goffin, C., & Ansari, D. (2016). Beyond magnitude: Judging ordinality of symbolic number is unrelated to magnitude comparison and independently relates to individual differences in arithmetic. *Cognition*, *150*, 68–76. https://doi.org/10.1016/j.cognition.2016.01.018

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini Meta-Analysis of Your Own Studies: Some Arguments on Why and a Primer on How: Mini Meta-Analysis. *Social and Personality Psychology Compass*, *10*(10), 535–549. https://doi.org/10.1111/spc3.12267

Gordon, P. (2004). Numerical Cognition Without Words: Evidence from Amazonia. *Science*, *306*(5695), 496–499. https://doi.org/10.1126/science.1094492

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. https://doi.org/10.1038/nature07246

Hauser, M. D., & Spelke, E. (2004). Evolutionary and developmental foundations of human knowledge: a case study of mathematics. In M. S. Gazzaniga (Ed.), *Cognitive Neurosciences* (3rd ed., pp. 853–864). Cambridge, MA: MIT Press.

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

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology*, *103*(1), 17–29. https://doi.org/10.1016/j.jecp.2008.04.001

IBM Corp. (2012). IBM SPSS Statistics for Windows (Version 21.0). Retrieved from https://www-01.ibm.com/support/docview.wss?uid=swg24032236

Ischebeck, A., Heim, S., Siedentopf, C., Zamarian, L., Schocke, M., Kremser, C., … Delazer, M. (2008). Are numbers special? Comparing the generation of verbal materials from ordered categories (months) to numbers and other categories (animals) in an fMRI study. *Human Brain Mapping*, *29*(8), 894–909. https://doi.org/10.1002/hbm.20433

Ito, Y., & Hatta, T. (2003). Semantic processing of Arabic, Kanji, and Kana numbers: Evidence from interference in physical and numerical size judgments. *Memory & Cognition*, *31*(3), 360–368. https://doi.org/10.3758/BF03194394

Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences*, *106*(25), 10382–10385. https://doi.org/10.1073/pnas.0812142106

Jou, J. (2003). Multiple number and letter comparison: directionality and accessibility in numeric and alphabetic memories. *The American Journal of Psychology*, *116*(4), 543–579.

Kallai, A. Y., & Tzelgov, J. (2009). A generalized fraction: An entity smaller than one on the mental number line. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1845–1864. https://doi.org/10.1037/a0016892

Kingdom, F. A. A., & Prins, N. (2010). *Psychophysics: A practical introduction*. San Diego, CA, US: Elsevier Academic Press.

Koechlin, E., Naccache, L., Block, E., & Dehaene, S. (1999). Primed numbers: Exploring the modularity of numerical representations with masked and unmasked semantic priming. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1882–1905. https://doi.org/10.1037/0096-1523.25.6.1882

Kojouharova, P., & Krajcsi, A. (2018). The Indo-Arabic distance effect originates in the response statistics of the task. *Psychological Research*. https://doi.org/10.1007/s00426-018-1052-1

Krajcsi, A. (2014, October). *A new framework for understanding numbers*. Presented at the CDC seminar series, CEU, Budapest, Hungary. Retrieved from http://cognitivescience.ceu.edu/node/2957

Krajcsi, A. (2016). Numerical distance and size effects dissociate in Indo-Arabic number comparison. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-016-1175-6

Krajcsi, A. (2018). *CogStat – An automatic analysis statistical software*. Retrieved from https://www.cogstat.org

Krajcsi, A., & Kojouharova, P. (2017). Symbolic Numerical Distance Effect Does Not Reflect the Difference between Numbers. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.02013

Krajcsi, A., Kojouharova, P., & Lengyel, G. (in preparation). *Development of Understanding Zero*. https://doi.org/10.20944/preprints201708.0087.v1

Krajcsi, A., Lengyel, G., & Kojouharova, P. (2016). The Source of the Symbolic Numerical Distance and Size Effects. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01795

Krajcsi, A., Lengyel, G., & Kojouharova, P. (2018). Symbolic Number Comparison Is Not Processed by the Analog Number System: Different Symbolic and Non-symbolic Numerical Distance and Size Effects. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00124

Krajcsi, A., Lengyel, G., & Laczkó, Á. (2018). Interference Between Number Magnitude and Parity: Discrete Representation in Number Processing. *Experimental Psychology*, *65*(2), 71–83. https://doi.org/10.1027/1618-3169/a000394

Landy, D. H., Jones, E. L., & Hummel, J. E. (2008). Landy et al 2008 - Why Spatial-Numeric Associations Aren't Evidence for a Mental Number Line. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *30*, 357–362. Retrieved from https://escholarship.org/uc/item/56f5w2k1

Leibovich, T., & Ansari, D. (2016). The symbol-grounding problem in numerical cognition: A review of theory, evidence, and outstanding questions. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *70*(1), 12–23. https://doi.org/10.1037/cep0000070

Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2017). From "sense of number" to "sense of magnitude": The role of continuous magnitudes in numerical cognition. *Behavioral and Brain Sciences*, *40*. https://doi.org/10.1017/S0140525X16000960

Leth-Steensen, C., Lucas, J., & Petrusic, W. M. (2011). Modelling SNARC by using polarity codes to adjust drift rates. *Proceedings of Fechner Day*, *27*, 357–362. Herzliya, Israel.

Leth-Steensen, C., & Marley, A. A. J. (2000). A model of response time effects in symbolic comparison. *Psychological Review*, *107*(1), 62–100. https://doi.org/10.1037/0033-295X.107.1.162

LibreOffice. (2018). *Home | LibreOffice - Free Office Suite - Fun Project - Fantastic People*. Retrieved from https://www.libreoffice.org/

Lipton, J. S., & Spelke, E. S. (2003). Origins of Number Sense: Large-Number Discrimination in Human Infants. *Psychological Science*, *14*(5), 396–401. https://doi.org/10.1111/1467-9280.01453

Lyons, Ian M., Ansari, D., & Beilock, S. L. (2012). Symbolic estrangement: Evidence against a strong association between numerical symbols and the quantities they represent. *Journal of Experimental Psychology: General*, *141*(4), 635–641. https://doi.org/10.1037/a0027248

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

Lyons, Ian M., & Beilock, S. L. (2013). Ordinality and the Nature of Symbolic Numbers. *The Journal of Neuroscience*, *33*(43), 17052–17061. https://doi.org/10.1523/JNEUROSCI.1775-13.2013

Lyons, Ian M., Nuerk, H.-C., & Ansari, D. (2015). Rethinking the implications of numerical ratio effects for understanding the development of representational precision and numerical processing across formats. *Journal of Experimental Psychology: General*, *144*(5), 1021–1035. https://doi.org/10.1037/xge0000094

Lyons, I.M., Vogel, S. E., & Ansari, D. (2016). On the ordinality of numbers. In *Progress in Brain Research* (Vol. 227, pp. 187–221). https://doi.org/10.1016/bs.pbr.2016.04.010

Maner, J. K. (2014). Let's Put Our Money Where Our Mouth Is: If Authors Are to Change Their Ways, Reviewers (and Editors) Must Change With Them. *Perspectives on Psychological Science*, *9*(3), 343–351. https://doi.org/10.1177/1745691614528215

Marinova, M., Sasanguie, D., & Reynvoet, B. (2018). Symbolic estrangement or symbolic integration of numerals with quantities: Methodological pitfalls and a possible solution. *PLOS ONE*, *13*(7), e0200808. https://doi.org/10.1371/journal.pone.0200808

Marshuetz, C., Reuter-Lorenz, P. A., Smith, E. E., Jonides, J., & Noll, D. C. (2006). Working memory for order and the parietal cortex: An event-related functional magnetic resonance imaging study. *Neuroscience*, *139*(1), 311–316. https://doi.org/10.1016/j.neuroscience.2005.04.071

Matejko, A. A., & Ansari, D. (2016). Trajectories of Symbolic and Nonsymbolic Magnitude Processing in the First Year of Formal Schooling. *PLOS ONE*, *11*(3), e0149863. https://doi.org/10.1371/journal.pone.0149863

Matejko, A. A., & Ansari, D. (2018). Contributions of functional Magnetic Resonance Imaging (fMRI) to the study of numerical cognition. *Journal of Numerical Cognition*, *4*(3), 505–525. https://doi.org/10.5964/jnc.v4i3.136

Matejko, A. A., Hutchison, J. E., & Ansari, D. (2018). Developmental specialization of the left intraparietal sulcus for symbolic ordinal processing. *Cortex*. https://doi.org/10.1016/j.cortex.2018.11.027

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), 320–334. https://doi.org/10.1037/0097-7403.9.3.320

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

Mussolin, C., Nys, J., Content, A., & Leybaert, J. (2014). Symbolic Number Abilities Predict Later Approximate Number System Acuity in Preschool Children. *PLoS ONE*, *9*(3), e91839. https://doi.org/10.1371/journal.pone.0091839

Nieder, A. (2005). Counting on neurons: the neurobiology of numerical competence. *Nature Reviews Neuroscience*, *6*(3), 177–190. https://doi.org/10.1038/nrn1626

Noël, M.-P., & Rousselle, L. (2011). Developmental Changes in the Profiles of Dyscalculia: An Explanation Based on a Double Exact-and-Approximate Number Representation Model. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00165

Nuerk, H., Iversen, W., & Willmes, K. (2004). Notational Modulation of the SNARC and the MARC (Linguistic Markedness of Response Codes) Effect. *The Quarterly Journal of Experimental Psychology Section A*, *57*(5), 835–863. https://doi.org/10.1080/02724980343000512

Núñez, R. E. (2017). Is There Really an Evolved Capacity for Number? *Trends in Cognitive Sciences*, *21*(6), 409–424. https://doi.org/10.1016/j.tics.2017.03.005

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716–aac4716. https://doi.org/10.1126/science.aac4716

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, *5*(5), 1. https://doi.org/10.1167/5.5.1

Patro, K., Nuerk, H.-C., Cress, U., & Haman, M. (2014). How number-space relationships are assessed before formal schooling: A taxonomy proposal. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00419

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

Piazza, M., Giacomini, E., Le Bihan, D., & Dehaene, S. (2003). Single-trial classification of parallel pre-attentive and serial attentive processes using functional magnetic resonance imaging. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1521), 1237–1245. https://doi.org/10.1098/rspb.2003.2356

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning Curves for Approximate Numerosity in the Human Intraparietal Sulcus. *Neuron*, *44*(3), 547–555. https://doi.org/10.1016/j.neuron.2004.10.014

Piazza, M., Mechelli, A., Butterworth, B., & Price, C. J. (2002). Are Subitizing and Counting Implemented as Separate or Functionally Overlapping Processes? *NeuroImage*, *15*(2), 435–446. https://doi.org/10.1006/nimg.2001.0980

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A Magnitude Code Common to Numerosities and Number Symbols in Human Intraparietal Cortex. *Neuron*, *53*(2), 293–305. https://doi.org/10.1016/j.neuron.2006.11.022

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science (New York, N.Y.)*, *306*(5695), 499–503. https://doi.org/10.1126/science.1102085

Pinhas, M., & Tzelgov, J. (2012). Expanding on the mental number line: Zero is perceived as the “smallest”. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1187–1205. https://doi.org/10.1037/a0027390

Platt, J. R., & Johnson, D. M. (1971). Localization of position within a homogeneous behavior chain: Effects of error contingencies. *Learning and Motivation*, *2*(4), 386–414. https://doi.org/10.1016/0023-9690(71)90020-8

Potts, G. R. (1972). Information processing strategies used in the encoding of linear orderings. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 727–740. https://doi.org/10.1016/S0022-5371(72)80007-0

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin*, *132*(3), 416–442. https://doi.org/10.1037/0033-2909.132.3.416

Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, *9*(3), 438–481. https://doi.org/10.3758/BF03196302

Razpurker-Apfeld, I., & Koriat, A. (2006). Flexible mental processes in numerical size judgments: The case of hebrew letters that are used to convey numbers. *Psychonomic Bulletin & Review*, *13*(1), 78–83. https://doi.org/10.3758/BF03193816

Reynvoet, B., & Brysbaert, M. (1999). Single-digit and two-digit Arabic numerals address the same semantic number line. *Cognition*, *72*(2), 191–201. https://doi.org/10.1016/S0010-0277(99)00048-7

Reynvoet, B., De Smedt, B., & Van den Bussche, E. (2009). Children's representation of symbolic magnitude: The development of the priming distance effect. *Journal of Experimental Child Psychology*, *103*(4), 480–489. https://doi.org/10.1016/j.jecp.2009.01.007

Reynvoet, B., & Sasanguie, D. (2016). The Symbol Grounding Problem Revisited: A Thorough Evaluation of the ANS Mapping Account and the Proposal of an Alternative Account Based on Symbol–Symbol Associations. *Frontiers in Psychology*, *07*. https://doi.org/10.3389/fpsyg.2016.01581

Sasanguie, D., De Smedt, B., & Reynvoet, B. (2017). Evidence for distinct magnitude systems for symbolic and non-symbolic number. *Psychological Research*, *81*(1), 231–242. https://doi.org/10.1007/s00426-015-0734-1

Sasanguie, D., Defever, E., Maertens, B., & Reynvoet, B. (2014). The Approximate Number System is not Predictive for Symbolic Number Processing in Kindergarteners. *Quarterly Journal of Experimental Psychology*, *67*(2), 271–280. https://doi.org/10.1080/17470218.2013.803581

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*(3), 418–431. https://doi.org/10.1016/j.jecp.2012.10.012

Sasanguie, D., Lyons, I. M., De Smedt, B., & Reynvoet, B. (2017). Unpacking symbolic number comparison and its relation with arithmetic in adults. *Cognition*, *165*, 26–38. https://doi.org/10.1016/j.cognition.2017.04.007

Sathian, K., Simon, T. J., Peterson, S., Patel, G. A., Hoffman, J. M., & Grafton, S. T. (1999). Neural Evidence Linking Visual Object Enumeration and Attention. *Journal of Cognitive Neuroscience*, *11*(1), 36–51. https://doi.org/10.1162/089892999563238

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Developmental Science*, *20*(3), e12372. https://doi.org/10.1111/desc.12372

Scholz, K. W., & Potts, G. R. (1974). Cognitive processing of linear orderings. *Journal of Experimental Psychology*, *102*(2), 323–326. https://doi.org/10.1037/h0035873

Schwenk, C., Sasanguie, D., Kuhn, J.-T., Kempe, S., Doebler, P., & Holling, H. (2017). (Non-)symbolic magnitude processing in children with mathematical difficulties: A meta-analysis. *Research in Developmental Disabilities*, *64*, 152–167. https://doi.org/10.1016/j.ridd.2017.03.003

Sekuler, R., & Mierkiewicz, D. (1977). Children's Judgments of Numerical Inequality. *Child Development*, *48*(2), 630. https://doi.org/10.2307/1128664

Seymour, P. H. K. (1980). Internal representation of the months: An experimental analysis of spatial forms. *Psychological Research*, *42*(3), 255–273. https://doi.org/10.1007/BF00308532

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*(3), 161–168. https://doi.org/10.1016/j.tins.2004.01.006

Stoianov, I., & Zorzi, M. (2012). Emergence of a "visual number sense" in hierarchical generative models. *Nature Neuroscience*, *15*(2), 194–196. https://doi.org/10.1038/nn.2996

Takahashi, A., & Green, D. (1983). Numerical judgments with Kanji and Kana. *Neuropsychologia*, *21*(3), 259–263. https://doi.org/10.1016/0028-3932(83)90042-8

Turconi, E., Campbell, J. I. D., & Seron, X. (2006). Numerical order and quantity processing in number comparison. *Cognition*, *98*(3), 273–285. https://doi.org/10.1016/j.cognition.2004.12.002

Tzelgov, J., Ganor-Stern, D., & Maymon-Schreiber, K. (2009). The Representation of Negative Numbers: Exploring the Effects of Mode of Processing and Notation.

Quarterly Journal of Experimental Psychology, *62*(3), 605–624. https://doi.org/10.1080/17470210802034751

Tzelgov, J., Yehene, V., Kotler, L., & Alon, A. (2000). Automatic comparisons of artificial digits never compared: Learning linear ordering relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 103–120. https://doi.org/10.1037/0278-7393.26.1.103

van Hoogmoed, A. H., & Kroesbergen, E. H. (2018). On the Difference Between Numerosity Processing and Number Processing. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01650

Van Opstal, F., Gevers, W., De Moor, W., & Verguts, T. (2008). Dissecting the symbolic distance effect: Comparison and priming effects in numerical and nonnumerical orders. *Psychonomic Bulletin & Review*, *15*(2), 419–425. https://doi.org/10.3758/PBR.15.2.419

Verguts, T., & Fias, W. (2004). Representation of Number in Animals and Humans: A Neural Model. *Journal of Cognitive Neuroscience*, *16*(9), 1493–1504. https://doi.org/10.1162/0898929042568497

Verguts, T., Fias, W., & Stevens, M. (2005). A model of exact small-number representation. *Psychonomic Bulletin & Review*, *12*(1), 66–80. https://doi.org/10.3758/BF03196349

Verguts, T., & Van Opstal, F. (2014). A delta-rule model of numerical and non-numerical order processing. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 1092–1102. https://doi.org/10.1037/a0035114

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, *36*(3), 1–48.

Vigliocco, G., Vinson, D. P., Damian, M. F., & Levelt, W. (2002). Semantic distance effects on object and action naming. *Cognition*, *85*(3), B61–B69. https://doi.org/10.1016/S0010-0277(02)00107-5

Wagenmakers, E.-J., Van Der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*(1), 3–22. https://doi.org/10.3758/BF03194023

Wallace, B. C., Danabreh, I. J., Trikalinos, T. A., Lau, J., Trow, P., & Schmid, C. H. (2012). Closing the gap between methodologists and end-users: R as a computational back-end. *Journal of Statistical Software*, *49*(5), 1–15.

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

Welford, A. T. (1960). The measurement of sensory-motor performance : Survey and reappraisal of twelve years' progress. *Ergonomics*, *3*(3), 189–230. https://doi.org/10.1080/00140136008930484

Wynn, K. (1998). Psychological foundations of number: numerical competence in human infants. *Trends in Cognitive Sciences*, *2*(8), 296–303. https://doi.org/10.1016/S1364-6613(98)01203-0

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1–B11. https://doi.org/10.1016/S0010-0277(99)00066-9

Zbrodoff, N. J., & Logan, G. D. (2005). What everyone finds: The problem-size effect. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (Psychology Press, pp. 331–345). New York, NY, US.

Zorzi, M., Di Bono, M. G., & Fias, W. (2011). Distinct representations of numerical and non-numerical order in the human intraparietal sulcus revealed by multivariate pattern recognition. *NeuroImage*, *56*(2), 674–680. https://doi.org/10.1016/j.neuroimage.2010.06.035

To my life partner

for always being there for me

To my parents

for their support

To my supervisor

for not giving up on me

To the Starman

To the Lightbringer

To L'Étranger

for my still being here