Hunting for the SNARC:
Spatial-Numerical Associations in an Order Judgment Task are Surprisingly Elusive

by
Dean Verger

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial
fulfillment of the requirements for the degree of

PhD
in
Psychology

Department of Psychology
Carleton University
Ottawa, Ontario

© 2015

Dean Verger

Abstract

Spatial-numerical associations (SNA) have been studied in various numerical and non-numerical tasks. Classically, participants respond faster with their right hand to larger numbers and faster with their left hand to smaller numbers even when number identity is irrelevant. SNAs are often explained as activation of a left-to-right mental number line. Thus, SNAs may reflect magnitude processes, relative order, or both. In the present research, I explored the SNA in an order judgment task. In three experiments, participants judged whether three digit sequences were ordered (e.g., 1 2 3) or unordered (e.g., 2 1 3). People with more efficient calculation skill are faster and more accurate on order judgments than those with less efficient skills. Thus this task is assumed to capture important aspects of the mental representation of numerical sequences. Accordingly, I hypothesized that participants would show a SNA similar to that in other number tasks (e.g., parity, number comparison), that is, faster right hand responses to number sequences that are assumed to activate the right side (e.g., 6 7 8) and faster left hand responses to number sequences that are assumed to activate the left side (e.g., 2 3 4) of the mental number line. In three experiments, ordered but not unordered sequences showed consistent evidence for a typical SNA. In Experiment 1, Asian-educated individuals showed a typical, stable SNA effect on both ascending and descending sequences whereas Arabic-speaking participants showed no significant SNA effects. English-speaking participants showed a typical SNA on descending sequences and a reversed SNA on ascending sequences. In Experiments 2 and 3, only individuals with better arithmetic fluency showed a typical SNA. In Experiments 2 and 3, analyses using a multi-factor ANOVA design did not show evidence for SNARC effects. Instead, SNARC

effects were found only using slope analyses on differences in responses with the right versus the left hand. The results are discussed in relation to the processes involved in sequence judgments and how those might differ from the processes involved in other numerical tasks. In summary, the SNA was surprisingly elusive, even though participants were quite sensitive to number magnitude and the task itself required order judgments, which presumably should have activated the mental number line.

Acknowledgements

Some years ago, I set off on a journey. Along the way there were those who provided guidance, when I was willing to look and listen; those whose laughter coloured a grey day; those whose own adventures lent wind to my sails. These, then, are my thanks to all who made the travel through ideas so worthwhile.

Many years ago I encountered Dr William Petrusic. He taught me to follow the data. He inspired me to pursue my PhD. He was a rich repository of knowledge, of stories, and of a belief in a just world. Thank-you to Elaine Petrusic for the many cups of tea, cookies and conversation whenever I would come into their home to talk about research. Bill's passing was a loss.

I could not have completed this degree without the support of my supervisor, Dr. Jo-Anne LeFevre. Jo-Anne provided me with an intellectual home in her lab. She encouraged me to explore more deeply both critical thinking and academic writing. She took the extra time to learn about my area of research, and encouraged me to think about alternative interpretations of my findings.

Thank-you to the folks in the psychophysics lab: Jordan Schoenherr and Steve Carroll. And to those in the math lab, my second home: Carla Sowinski, Heather Douglas, Claire Xu, Chunyun Ma, Carolina Jimenez, and Matthew Huebner.

Thank-you to June Callender and Etelle Bourassa, who always answered my questions, and ably plucked me out of the briar patches into which I repeatedly fell.

And finally, thank-you to my co-adventurer, my life-partner Ruth Stewart-Verger, whose spirit steadies my own. She is my keel on this sea of life, as we joyously sail forward together.

Table of Contents

Abstract	ii
Acknowledgements	iv
Table of Contents	v
List of Tables	ix
List of Figures	xii
CHAPTER 1: Introduction	1
Number knowledge.....	1
Order and Magnitude.....	3
Spatial Numerical Association of Response Codes.....	4
Explanations of spatial-numerical associations	5
Two hypotheses about magnitude representations: The mental number line versus the numerosity code hypotheses	7
Overview of the Thesis.....	9
CHAPTER 2: Literature Review	11
The original SNARC experiments.....	11
Ordered sequences	15
Contextual Factors that May Influence SNARC Effects	21
Computational skills.....	21
Mathematics education	24
Number Language	25
Analyzing SNARC	30
Power Calculations	32

Summary of Hypotheses.....	33
CHAPTER 3 – EXPERIMENT 1	37
Method	38
Participants	38
Equipment.....	39
Materials	39
Procedure	41
Analysis.....	42
Data Trimming	42
Speed-Accuracy Trade-off	42
Error Analysis.....	42
Error Summary	48
ANOVAs on Mean Response Latencies	49
SNARC Analysis.....	55
Power Analysis	59
Discussion.....	59
Modifications for Experiment 2	63
CHAPTER 4 – EXPERIMENT 2	65
Method	67
Participants	67
Materials	67
Equipment.....	68
Procedure	69

Results	70
Data Trimming	70
Speed-Accuracy Trade-off	70
Calculation Fluency Test.....	71
Error Analysis.....	71
ANOVA on Mean Response Times	76
SNARC Effects.....	83
Power Analysis	90
Discussion	90
Limitations of Experiment 2.....	91
CHAPTER 5 – EXPERIMENT 3	93
Method	95
Participants	95
Equipment.....	96
Materials	97
Procedure	98
Results	100
Skill Groups.....	100
Sequence Task	101
Power analysis for Sequence task.....	113
Parity Task.....	114
Analysis for Parity task.....	122
Correlations between calculation fluency, and slopes for Parity and Sequence ..	123

Discussion	128
Limitations of Experiment 3	130
Chapter 6 – General Discussion.....	131
Summary of Findings	132
Discussion of Findings	134
Arithmetic Skill	137
Numerosity Code versus Mental Number Line.....	138
A Proposed Model.....	140
Limitations	143
Future Research.....	146
Conclusion.....	147
References	148
APPENDICES	158
Appendix A: Power Analyses	158
Appendix B – Extra Experiment 1 figures	161
Appendix C - Previous SNARC experiments	163

List of Tables

Table 2.1 Summary of thesis experiments and hypotheses	36
Table 3.1. Experiment 1: Summary information of first languages of participants, the direction of reading in the first language, and additional language fluencies.	39
Table 3.2. Experiment 1: Single-digit number sets (Triads).....	40
Table 3.3. Experiment 1: Source table for repeated measures ANOVA on percentage errors	44
Table 3.4. Experiment 1: Source table for the within-subjects ANOVA on mean response time.	50
Table 3.5. Experiment 1: Summary of the SNARC effect for the First-Language groups. Type is Stimulus type (ordered and unordered).....	57
Table 4.1. Experiment 2: Summary information of languages read by participants.	67
Table 4.2. Experiment 2: The single-digit number sets (Triads)	68
Table 4.3. Experiment 2: Table of descriptive statistics for Calculation Fluency Test scores.....	71
Table 4.4. Experiment 2: Source table of repeated measures ANOVA for percentage errors.	73
Table 4.5. Experiment 2: Source table for the within-subject ANOVA on mean response time.	78
Table 4.6. Experiment 2: Summary of the SNARC effect	84
Table 4.7. Experiment 2: Summary of the SNARC effect and the effects of Skill.	85
Table 4.8. Experiment 2: Summary of the comparison of SNARC effects between low and high Arithmetic skill groups for Task and Stimulus Type.	86

Table 4.9. Experiment 2: Correlations among SNARC coefficients and Calculation Fluency scores.....	89
Table 5.1. Experiment 3: Summary information of languages read by participants (in alphabetical order).	96
Table 5.2. Experiment 3: Single-digit number sets (Triads).....	98
Table 5.3. Experiment 3: Table of descriptive statistics for the Calculation Fluency Test.	101
Table 5.4. Experiment 3: Source table of repeated measures ANOVA for percentage errors.	103
Table 5.5. Experiment 3: Source table for the analysis of mean response times in the Sequence Judgment Task.....	107
Table 5.6. Summary of the ordered Stimulus type SNARC effects by Skill group.	111
Table 5.7. Experiment 3: Summary of the unordered Stimulus type SNARC effects by Training Group.....	112
Table 5.8. Experiment 3: Summary of the unordered Stimulus type SNARC effects for the control group.	112
Table 5.9. Experiment 3: Source of repeated measures ANOVA for percentage of errors.	116
Table 5.10. Experiment 3: Source for the within-subject ANOVA on mean response times for the Parity Judgment Task.	118
Table 5.11. Experiment 3: Summary of single-sample <i>t</i> -tests on the slopes for the Parity Judgment Task.	121

Table 5.12. Experiment 3: Correlations among Calculation fluency scores and slopes for the Parity and Sequence tasks ($n = 71$).	123
Table 5.13. Experiment 3: Correlation table between Calculation fluency scores and individual slopes for the Parity and Sequence tasks ($n = 70$, outlier removed).....	127
Table A.1. Summary of power analysis for Experiment 1.	158
Table A.2. Summary of power analysis for Experiment 2.	159
Table A.3. Summary of power analysis for ordered sequences Experiment 3.	159
Table A.4. Summary of power analysis for unordered sequences for Experiment 3.	160
Table A.5. Summary of power analysis for unordered sequences for controls for Experiment 3.	160
Table A.6 Summary of power analysis for parity for Experiment 3.	160
Table C.1 Summary of experiments using non-numeric stimuli and the factors tested listed by date	163
Table C.2 Summary of experiments involving synaesthetes and the factors tested listed by date	163
Table C.3 Summary of experiments using numbers and the factors tested listed by date	164

List of Figures

Figure 2.1. A fictional example scatterplot of difference RTs under a parity judgment task. The dotted line is the line of best fit.	11
Figure 3.1. Experiment 1: Interaction plot of Task by Stimulus type on percentage errors. Error bars are 95% CIs.	45
Figure 3.2. Experiment 1: Interaction plot of Stimulus type by Triad by Language on percentage errors. Error bars are 95% CIs.	46
Figure 3.3. Experiment 1: Interaction plot of Task by Triad by Language on (right – left hand) percentage errors . Positive values indicate more errors with the right hand than the left. Error bars are 95% CIs.	47
Figure 3.4. Experiment 1: Line plot of Response Times for each Triad. The error bars are 95% CIs.	51
Figure 3.5. Experiment 1: Interaction plot of Response Times for Stimulus Type by Triad. The error bars are CIs 95%.	52
Figure 3.6. Experiment 1: Interaction plot of Response Times for Triad by Sequence type by Side of response for the unordered stimuli. The error bars are 95%CIs.	54
Figure 3.7. Experiment 1: Line plots of the SNARC effects for ordered stimuli. The lines are the lines of best fit.	59
Figure 4.1. Experiment 2: Interaction plot of Task by Triad for percentage of errors. The error bars are 95% CIs.	72
Figure 4.2. Experiment 2: Interaction plot of Task by Stimulus Type by Triad for percentage errors. The error bars are 95% CIs.	75

Figure 4.3. Experiment 2: Interaction plot of Stimulus Type by Triad for percentage errors (right – left hand). The error bars are 95% CIs.	76
Figure 4.4. Experiment 2: Line plot of mean response time for Triads. The error bars are 95%CIs.....	79
Figure 4.5. Experiment 2: Interaction plot of mean response time for Task Direction by Stimulus type of stimulus set. The error bars are 95%CIs	80
Figure 4.6. Experiment 2: Interaction plot of mean response time for the interaction of Task by Triad. The error bars are 95% CIs.....	80
Figure 4.7. Experiment 2: Interaction plot of mean response times for Task by Stimulus type by Triads. The error bars are 95%CIs.	82
Figure 4.9. Experiment 2: Scatterplot of slopes for descending unordered and ascending ordered. Line is line of best fit ($n = 26$).	90
Figure 5.1. Experiment 3: Interaction plot of Stimulus Type by Triad for mean percentage errors. The error bars are 95% CIs.	104
Figure 5.10. Experiment 3: Interaction plot of Number by Side of Response for mean Response Times on Parity Judgments. Error bars are 95%CIs.....	120
Figure 5.11. Experiment 3: Scatterplot of the difference in Response Times for the Parity task, for the Odd numbers (black circles) and the Even numbers (orange diamonds). Lines of best fit are shown.	122
Figure 5.12. Experiment 3: Scatterplot of slopes for ordered Sequences by Math score. Line is line of best fit ($n = 71$).	124
Figure 5.13. Experiment 3: Scatterplot of slopes for unordered Sequences by slopes for Parity task. Line is line of best fit ($n = 71$).	125

Figure 5.14. Experiment 3: Scatterplot of slopes for ordered Sequences by slopes for Parity task showing best fitting line ($n = 71$).	126
Figure 5.15. Experiment 3: Scatterplot of slopes for unordered Sequences by slopes for Parity task. Line is line of best fit.	127
Figure 5.2. Experiment 3: Interaction plot of Stimulus Type by Arithmetic skill for mean percentage errors. The error bars are 95% CIs.	105
Figure 5.3. Experiment 3: Line plot of Triad by mean Response Time. The error bars are 95% CIs.	108
Figure 5.4. Experiment 3: Interaction plot of Arithmetic skill by Triad by mean Response Time. The error bars are 95% CIs.	108
Figure 5.5. Experiment 3: Interaction plot of Stimulus type by Triad for mean Response Time. The error bars are 95%CIs.	109
Figure 5.6. Experiment 3: Scatterplot of ordered Triads by difference in mean response time by Arithmetic skill.	111
Figure 5.7. Experiment 3: Interaction plot of Numbers by Side of response (Yes = right, Yes = left) for Parity on mean percent errors. The error bars are 95% CIs.	117
Figure 5.8. Experiment 3: Mean response times by Number for Parity Judgments. Error bars are 95%CIs.	118
Figure 5.9. Experiment 3: Interaction plot for mean response times of Number by Arithmetic skill for Parity Judgments. Error bars are 95%CIs.	119
Figure 6.1. Model of judgment process.	143
Figure B.1. Interaction plot of Response Times for Triad by Side of response for the unordered stimuli for English group. The error bars are CI 95%	161

Figure B.2. Interaction plot of Response Times for Triad by Side of response for the ordered stimuli for English group. The error bars are CI 95%	161
Figure B.3. Interaction plot of Response Times for Triad by Side of response for the ordered stimuli for Arabic group. The error bars are CI 95%	162
Figure B.4. Interaction plot of Response Times for Triad by Side of response for the ordered stimuli for Arabic group. The error bars are CI 95%	162

CHAPTER 1: INTRODUCTION

When people judge whether a number is smaller or larger than another number, or decide if a number is odd or even, they respond faster to small numbers with their left hand and faster to large numbers with their right hand. This hand by side-of-response bias is known as the Spatial Numerical Association of Response Codes, or SNARC for short (Dehaene, Bossini & Giraux, 1993). What mechanism is at play? Magnitude was originally deemed the causal factor in SNARC, on the assumption that numbers are organized on a mental number line that represents the magnitude of numbers, and this mental number line starts at 1 (or 0) on the left and increases towards the right. However, other researchers have found a SNARC effect using non-numeric stimuli, such as letters of the alphabet, days of the week, or months of the year, stimuli that are ordered but do not have magnitude (Gevers, Reynvoet, & Fias, 2003; Gevers, Reynvoet, & Fias, 2004). As of yet, the role that order plays in the SNARC effect for numbers has not been directly tested. Accordingly, the purpose of this thesis was to examine the contribution of order to the numerical SNARC.

Number knowledge

As children develop number knowledge, they acquire an understanding of numerosity (i.e., magnitude), that is, the concept of ‘how many’, as well as an understanding of order, that is, that numbers are arranged in an invariant sequence. They begin by learning number names. Their number associations take shape, often with the help of their own fingers and toes (Fischer & Brugger, 2011; Fuson, 1988; Gelman & Gallistel, 1978). By age 5 children learn to associate arrays (groups of objects), digits (the symbolic representation of quantity) and number-words (Benoit, Lehalle, Molina, Tijus,

& Jouen, 2013). After learning the names of the numbers and their correct order, that is, “one, two, three”, understanding that two refers to any two items, and recognising that stopping the count at the third object indicates that there are three objects, children have learned to apply the principle of cardinality (Gelman & Gallistel, 1978; Mix, Sandhofer, & Baroody, 2005). Thus, children learn magnitude and order information simultaneously as they acquire early numeracy skills.

The inter-relation between order and magnitude continues beyond foundational aspects of counting and number. For example, early addition skills involve the count-on strategy, in which children identify the smaller of two numbers and count on from the larger, requiring them to make a magnitude judgment, for example, $2 + 5$ is best solved by counting on 6, 7 rather than 3, 4, 5, 6, 7 (Siegler & Shrager, 1984). As number skills improve, children shift away from working memory-intensive procedural use, such as counting, to more efficient and accurate direct recall (Bailey, Littlefield, & Geary, 2012). With time, our mental representation of numbers includes many aspects of number, including both order (e.g., 1 2 3) and magnitude (3 is larger than 2). When we make judgments about numbers we access both order and magnitude information. However, although order and magnitude are very closely connected, some theorists have suggested that these constructs are separable (Rubinsten & Sury, 2011).

Parity is another learned aspect of number. If a number is divisible by 2 it is even, else it is odd. Hence, parity requires knowledge of division, knowing if a number is a multiple of 2, or skip counting, beginning with either 1 or 2, resulting in two sequences, {1 3 5 7 9} and {2 4 6 8}. Parity does not require knowledge of quantity or order. With practice, the determination of parity can be done quickly and efficiently via recall from

long-term memory or with more time and effort by calculation. Parity is a number feature that is learned by Grade 2 (Berch, Foley, Hill, & Ryan, 1999). For elementary school students the main criteria for decision-making in similarity judgments remains magnitude. Parity only begins to appear as a factor in Grade 6 (Miller & Gelman, 1983). Learned in elementary school, we retain knowledge of parity, order, and magnitude into adulthood. In summary, our experience with numbers begins early on. We learn that a number is more than a name or a quantity. We learn to recognize the ordering of the number sequence 1 through 9, that numbers farther along the sequence are greater than the early numbers, and that each number may have a special attribute, such as being odd or even.

Order and Magnitude

The use of quantity and order are intertwined. Numbers that are different in magnitude are also separated sequentially, so a larger size difference corresponds to a greater ordinal distance. Despite the close connection between magnitude and order, theorists have suggested that order and magnitude are separable constructs (Rubinsten & Sury, 2011). Empirical work has supported this view. For example, Zhao et al. (2012) found a neural dissociation between magnitude and order when people learned these features separately. Nieder and Miller (2004) showed in brain experiments with monkeys that magnitudes do not need to be associated with order, as Zhao et al. (2012) did with humans. Thus, processing of magnitude and order might be “based on different mechanisms” (Badets, Boutin, & Heuer, 2015, p. 173). Similarly, Turconi, Campbell, and Seron (2006) concluded that “...quantity and order judgments recruited distinct cognitive mechanisms” (p. 282). Accordingly, magnitudes can be directly tied to quantity (Zorzi,

Di Bono, & Fias, 2011) and thus need not necessarily activate relative order. The view that magnitudes can be represented separately from order was termed the numerosity code hypothesis by Zhao et al. (2012), and contrasted to the mental number line hypothesis, in which order and quantity are necessarily linked.

Spatial Numerical Association of Response Codes

The study of magnitude in simple number tasks has been influenced very strongly by the observation of the SNARC effect, that is, the spatial-numerical association that occurs when people make responses to numerical stimuli. Dehaene et al. (1993) identified and named the SNARC effect in data collected from a parity judgment task on single-digit numbers. They attempted to determine whether the effect was specific to number by running a comparable experiment using letters. The assumption was that numbers include both order and magnitude within long-term memory representations whereas letters only included order information. Hence, if the SNARC effect was found for both numbers and letters, then it could be attributed to the shared factor of order. Because no hand-speed bias was found for letters, however, Dehaene et al. (1993) concluded that the effect must have been due to magnitude, not order.

Other researchers improved the experimental design, and found a SNARC effect using stimuli such as months of the year, days of the week, seating arrangements, sequences of unrelated words, and unknown heights of men (Gevers et al., 2003; Gevers et al., 2004; Prado, Van der Henst, & Noveck, 2008; Previtali, de Hevia, & Girelli, 2010; Shaki, & Leth-Steensen, 2009). In summary, contrary to Dehaene et al.'s (1993) conclusion, order appears to unite the SNARC for letters, months, heights and other sequentially related stimuli, including sets of non-numeric arbitrary symbols. Despite

these findings, the SNARC effect is still typically assumed to index magnitude codes in numerical tasks and is often cited as evidence for a mental number line (reviewed by Fias & Fischer, 2005; Hubbard, Piazza, Pinel, & Dehaene, 2005).

Explanations of spatial-numerical associations

Nevertheless, the source of the SNARC effect remains somewhat controversial and thus there are different theories about its source, such as the reading instruction account (Dehaene et al., 1993; Göbel, Shaki, & Fischer, 2011; Zebian, 2005), the strategic view (Fischer, 2006; van Dijck & Fias, 2011), the brain lateralization account (de Hevia et al., 2014), and the implicit instruction account (Patro, Fischer, Nuerk, & Cress, 2015). Furthermore, some researchers (e.g., Dodd, Van der Stigchel, Leghari, Fung, & Kingstone, 2008) have proposed that the SNARC effect in numerical tasks is special, because numbers inherently carry both magnitude and ordinal information.

Reading (and writing) directions are culturally taught and learned. Years are spent developing skills that activate both the perceptual and the motor systems. Dehaene et al. (1993) noted that the direction of the SNARC slope for Arabic students who read right-to-left in their first language appeared to depend on how long they had been living in France. The longer their stay, the more French they learned, with its reading direction of left-to-right, the more the slope trended towards the negative. Other researchers have also found a correspondence between direction of reading and the SNARC effect (Göbel, 2011; Shaki, Petrusic, & Leth-Steensen, 2012; Zebian, 2005).

Not all spatial-numerical associations display an effect in the same direction as reading. Ito and Hatta (2004) found a vertical SNARC effect. The Japanese read from the top to bottom, then left to right in columns of script. However, the effect was bottom to

top rather than top to bottom. The reversal of the vertical SNARC effect may have been influenced by the physical experience. For example, quantity increases as the contents rise in a glass or a cup. Fischer (2006) proposed that, rather than automatically accessing a mental representation of each number, people can make strategic decisions based on the overall demands of the task. This explanation would explain why associations between number and space sometimes do not appear.

Contrary to the stance that SNARC arises out of learning to read, and the inherent spatial orientation, research with 7-month-old infants using a habituation technique suggests that people have a natural left-to-right predisposition (de Hevia et al., 2014).

Although de Hevia et al. concluded that people are born with a default left-to-right gaze, Patro et al. (2015) found that instruction could “shape number-space associations”. Children 3 to 4 years of age played a game. The task was to move a small green frog across a touch screen. The manipulation was the direction the frog needed to be moved to win. One half of the participants moved the frog from left-to-right whereas the other half of the children moved the frog from right-to-left. After only 15 trials, children who moved the frog towards the left demonstrated a leftward bias in a numerosity comparison task, whereas children who moved the frog toward the right demonstrated a rightward bias. This study suggests that spatial-numerical associations can be constructed with a minimal amount of experience.

Lastly, Dodd et al. (2008) described numbers as special after testing for the SNARC effect using a variety of stimuli (numbers, letters, days, and months). SNARC was found for numbers in the first experiment but not for the non-numeric stimuli. However, when order was included as a second task (i.e., is the stimulus before or after a

standard) then the SNARC effect was found for all stimuli. Accordingly, Dodd et al. argued that numbers do not need special instructions or contexts for a SNARC effect to arise and thus that order and magnitude are inextricably linked in number representations.

In summary, these different accounts of the source of the SNARC effect suggest that the SNARC effect is flexible. The reading instruction and the implicit instruction accounts argue that spatial-numerical associations are learned through our interactions with the culture. Even the brain lateralization account only indicates a starting point, and that only tested in one cultural context. The strategic view allows for the brain mechanisms to switch procedures as tasks and task demands change. In the present research I am presenting two similar tasks, judging ascending, and judging descending sequences. The participants will have experienced both in school, but may not have had recent experience with them. Judging the more difficult descending sequences may cause procedural shifts in how the sequences are judged.

Two hypotheses about magnitude representations: The mental number line versus the numerosity code hypotheses

Mental number line. Moyer and Landauer (1967) proposed that we maintain an analogue representation of the number line in long-term memory. They extended psychophysical research on absolute judgment experiments with single dimension stimuli such as line lengths and pitch to number comparisons. In physical judgments, effects of ratio and distance are observed, such that the difficulty of the decision increases as the ratio and distance between the stimuli decreases. Moyer and Landauer observed similar patterns in comparisons of single-digit numbers, that is, which is larger, 4 or 9? Accordingly, participants were faster and more accurate to judge relative magnitude for

numbers that were farther apart (4 and 8) than for numbers that were closer together (5 and 6). This pattern of results for numbers looked very much like the pattern found for the judgments about physical stimuli. Accordingly, Moyer and Landauer proposed that people translate numbers into their analogue magnitudes, and then make the comparison. This process is assumed to implicate a mental representation that is commonly referred to as the **mental number line** (a term that seems to have been first used by Aiken and Williams in 1973, though Restle (1970) did describe the use of a number line as an imaginary analog).

The hand-speed side of response bias known as SNARC was also explained in terms of a mental number line (Dehaene et al., 1993). In their experiments, the numbers that were used as stimuli appeared in the center of the monitor. Accordingly, the hand-speed differences of the correct responses that they observed were presumably related to the spatial location of each number in a mental representation, with small numbers being advantaged to the left, and large numbers to the right. Although neither magnitude nor order information was required for the parity judgment, the pattern of responses suggested that participants automatically activated number knowledge, including the location of each single-digit number relative to each other, when they made the parity decision. As noted earlier, Dehaene et al. concluded that the arrangement of the mental number line was an index of magnitude rather than of order, a conclusion that has been challenged in the intervening years (Gevers et al., 2003; Previtali et al., 2010).

The numerosity code hypothesis states that magnitudes are simply represented as quantities (Zhao et al, 2012; Zorzi, Stoianov, & Umiltà, 2005). If magnitudes are simply quantities, then order is a separate concept. In contrast, the number line hypothesis states

that numbers have an ordinal representation on a mental number line (Zhao et al., 2012). If numbers are represented with magnitude and order information as an integral part of each number, then it should always be activated in any task where numbers are activated, even when for tasks that do not require access to magnitude.

In summary, according to the mental number line hypothesis, any task involving judgments about numbers, even those that do not involve magnitude such as parity or order, should activate magnitude codes. The numerosity code hypothesis, in contrast, allows for number activations that involve magnitude to be separate from those that involve order, and so number tasks that only involve order could, at least hypothetically, not result in automatic activation of magnitude. This possibility was explored in the present thesis.

Overview of the Thesis

This thesis consists of six chapters. Chapter 2 includes an overview of the literature on the SNARC effect. Chapters 3 – 5 describe three experiments in which I explored the SNARC effect in a numerical task that is focused on order rather than magnitude. Chapter 6 is a summary of my empirical work and includes conclusions and discussion of future research. As noted earlier, my goal in this research was to examine the contribution of order to the numerical SNARC. Previous research has shown that SNARC effects occur both in numerical and non-numerical sequences. However, such findings do not prove that the SNARC effect in numerical tasks is due solely to order information. I hypothesized that if the SNARC effect is also related to magnitude in numerical tasks, then I should observe a magnitude-based SNARC effect in a numerical task that requires only order judgments. Conversely, if SNARC effects are primarily due

to order and not to magnitude processes then activating the order information in a symbolic number task may preclude SNARC effects that are due to magnitude. SNARC effects are commonly found in number tasks that required magnitude information, such as comparison, as well as in number tasks that apparently do not require magnitude information such as parity judgments. However, no research has explored SNARC effects in numerical tasks that are primarily about order (cf. Turconi et al., 2006). The results of the present research provide additional insights into the relation between order and magnitude information in numerical cognition.

CHAPTER 2: LITERATURE REVIEW

The original SNARC experiments

Dehaene et al. (1993) instructed university students to judge parity in a set of nine experiments. Experiment 1 used the Arabic numerals 0 – 9. In Experiment 1, the researchers found an intriguing space-number hand-speed difference: The left hand was faster than the right when responding to the relatively small numbers whereas, the right hand was faster than the left when responding to the relatively large numbers. Dehaene et al. labelled this effect the *spatial-numerical association of response codes*, or SNARC. The analyses are done on the difference scores (right hand RT – left hand RT) for each digit. Because right hand responses take more time than left hand responses, for the smaller numbers these differences are positive, whereas, the difference scores are negative for the larger numbers. This pattern of differences results in a negative slope. Figure 2.1 shows an example of a typical, but fictitious, SNARC graph for parity judgments.

Figure 2.1. An idealized example scatterplot of difference RTs under a parity judgment task. The dotted line is the line of best fit.

The researchers went on to examine the limits of the effect in eight additional experiments. In Experiment 4, Dehaene et al. (1993) used two different sets of letters as stimuli to test whether the SNARC effect reflected number magnitude or order (i.e., relative positions). The first task was made up of the first six letters of the alphabet assigned to two groups, A-C-E and B-D-F. Ten participants had to make a judgment on each trial about group membership, responding with the left hand for group 1 and the right hand for group 2 in one block and then reversing the hand of response for the second block. Another ten participants made decisions about a second set of letters. The letter stimuli were five consonants in group 1, and five vowels in group 2, in an alternating list (i.e., A C E G I L O R U) to more closely correspond to the alternating feature of even and odd for the number set 0 – 9. For both sets, the letters should have produced a SNARC effect if relative position influenced the judgment. However, no difference in hand speed was found across groups for either set of letters. As such, they concluded that the hand-speed difference for numbers was a magnitude effect, arising from the irrelevant intrusion of magnitude in the decision making process. The underlying assumption, therefore, was that the magnitude of the number resulted in automatic access to a mental representation of magnitude, arranged as a number line. This assumption, that SNARC effects reflected activation of magnitude as indexed by a mental number line received support from other researchers, including Miller and Gelman (1983) and Sudevan and Taylor (1987).

Gevers et al. (2003) were the first to address the issue of whether SNARC effects can be found in ordered stimuli other than numbers. They used the well-learned sequences of months of the year and letters of the alphabet as stimuli. They identified

three possible limitations of Experiment 4 in Dehaene et al. (1993) that might have explained the lack of SNARC for letters in the original letter experiment. First, Gevers et al. argued that the assignment of letters to two groups for the six chosen letters had been arbitrary and thus membership in the two letter groups would not have automatically accessed order from memory. In the original letter classification task the time to respond increased across the letter set {A, B, C, D, E, and F}. Gevers et al. interpreted this pattern as evidence for a serial search process rather than direct memory access, or automaticity. Second, Gevers et al. noted that the letter experiment that used alternating consonants and vowels had actually shown a trend that was consistent with a SNARC effect. But, with only ten participants, the lack of significance had likely been due to a lack of power. And thirdly, they stated that order had not been relevant to the task.

To address the limitations of the Dehaene et al. (1993) experiment, Gevers et al. (2003) more than doubled the number of participants to 25 and 24 in Experiments 1 and 2 respectively, and included both order-relevant and order-irrelevant comparison tasks. In their first experiment, the stimuli were the first four months and the last four months of year. In the order-relevant condition, participants judged whether the presented month came before or after July. In the order-irrelevant condition they judged whether the last letter of the month was “R” (the experiment was conducted in Dutch). In the second experiment they used nine letters, four vowels and four consonants with an additional midpoint vowel. The order-relevant condition required the participants to judge whether the letter came before or after the midpoint letter (O). The order-irrelevant condition required the participants to judge whether the letter was a vowel or a consonant. This procedure paralleled the parity task for numbers. SNARC effects were found in both

experiments. In a follow-up experiment using days of the week, the SNARC effect was again found for both the order-relevant and order-irrelevant conditions (Gevers et al., 2004).

Besides using well-known stimuli such as months and letters of the alphabet, other experiments have found significant SNARC effects using newly learned ordered sequences of unrelated words (Previtali et al., 2010), and order developed through relational reasoning of seating arrangements (Prado et al., 2008). Research measuring changes in the brain has also found significant SNARC effects for newly learned sequences of novel figures (Van Opstal, Fias, Peigneux, & Verguts, 2009). Together, these findings suggest that SNARC-like effects are commonly observed in non-numerical tasks that require judgments based on order, and thus suggest that the SNARC effect in numerical tasks may also reflect order, either instead of or in addition to, magnitude.

SNARC for numbers has been found in numerical tasks that explicitly require access to magnitude such as size comparisons (Bull, Cleland, & Mitchell, 2013; Fischer, 2004; Fischer & Rottmann, 2005; Fitousi, Shaki, & Algom, 2009; Notebaert, Gevers, Verguts, & Fias, 2006; Shaki & Petrusic, 2005), in attention tasks (Dodd et al., 2008; Fischer, Castel, Dodd & Pratt, 2003), in a task using recipes (Fischer, Mills & Shaki, 2010), and a task that replaced the usual left and right response assignments using two hands with a response that was closer or farther from a central button on a keyboard (Santens & Gevers, 2008). These findings are consistent with the initial position of Dehaene et al. (1993) that SNARC is a magnitude effect. However, because SNARC effects were found for non-numeric ordered sequences, and because order and magnitude are both represented in numerical information, the role of these two factors cannot be

separated easily. The goal of the present research was to determine if order played a role in the SNARC effect for numbers. Therefore, the next section reviews research that has been done on ordered number sequences.

Ordered sequences

The few SNARC studies that have used order judgments have done so with non-numeric stimuli. For example, Shaki and Leth-Steensen (2009) reported on their experiment, which used comparative judgments of learned relative heights. The stimuli were six three-letter names of men. During a learning phase, participants ($N = 45$) judged on half the trials which of two men was taller, and on the other half of the trials they judged which of two men was shorter. Participants were not informed as to the actual ordering of the heights of the six. Rather they learned via feedback. The participants were randomly assigned to four orderings of the six names. During the learning phase, only the five adjacent pairs were used. During the test phase all 15 possible pairwise comparisons were presented. However, analysis was only conducted on the adjacent pairings. A SNARC effect was found for the “choose shorter” instruction, and a reversed SNARC effect was found for the “choose taller”. Shaki and Leth-Steensen argued that the reversal of the SNARC effect demonstrated a contextual, or instruction-based SNARC. More important for the current focus on the role of ordinality in the SNARC, they also argued that because no actual heights had been detailed the relative heights of the stimuli as learned during the feedback phase was ordinal in scale, and did not activate magnitude. This study suggests that order and magnitude can be processed separately, at least in tasks where magnitude is not necessary for successful performance.

In one of the few studies which contrasted order and magnitude judgments with numerical stimuli, Turconi et al. (2006) studied differences in people's decisions about quantity and order in judgment tasks using pairs of digits. After conducting two experiments they concluded that quantity and order judgments tap into different processes. In Experiment 1 each participant performed two tasks. In one task, they made order judgments about whether two digits were in ascending order (e.g., 3 7 vs. 6 4). In the other task, they judged which of the two numbers was larger (e.g., 3 7 vs. 6 4). Turconi et al. were interested in the distance effect, both in its standard format and its reversed direction. Specifically, other research has shown that when people make magnitude decisions about two numbers far apart from each other, responses are quick. But, when the two numbers are close it takes longer to respond. This is the standard distance effect (Moyer & Landauer, 1967). In contrast, when participants make order judgments for ascending sequences, the response times for those that are close together (e.g., 3 4) are faster than for those sets of stimuli (e.g., 4 7) that are farther apart (Jou, 1997).

Turconi et al. (2006) used the same stimuli for the two tasks and manipulated distance, such that the difference between number pairs was either one (e.g., 2-3; close pairs) or three (e.g., 5-8; far pairs). They found that there was a main effect of task. Order judgments took longer than quantity judgments. Standard distance effects were found for both close and far pairs for the quantity task, and for unordered (i.e., descending) stimuli in the order task. In contrast, a reverse distance effect was found for ascending pairs with the response times being faster for the close pairs compared to the far pairs.

In Experiment 2 participants performed one of two tasks, either quantity comparison or order judgment. Two additional distances between number pairs were used in the construction of the stimuli for this experiment resulting in distances of 1, 2, 3, and 4. Again, there was a main effect of task. Order judgments were slower than quantity judgments. There were standard distance effects for quantity comparisons, and for order judgments of descending pairs, with responses to closer pairs taking more time than farther pairs. In contrast, there was both a standard and a reverse distance effect for the order judgments of ascending number pairs. Response times for Distance 1 and 4 were the same. Response times for both were faster than response times for Distance 3, which was faster than for Distance 2. The variability in the pattern of distance effects provides evidence that order judgment and quantity comparison use different mental processes.

Turconi et al. (2006) suggested two possible explanations for the reverse distance effect for order judgments. The first explanation is that participants defaulted to a serial search, which allowed relatively fast responses for Distance 1. However, when distance increased between number pairs, serial search would become inefficient. This might motivate a switch on the trial to the more efficient magnitude comparison procedure. The second explanation is that close sequences which could be recognized from memory because they form part of the very familiar counting sequence (e.g., 2 3, 4 5 and so on). Sequences of Distance 1 would then have a response time advantage. As distances increased, again order judgment would use a magnitude comparison and it would be faster as the distances were larger, producing the increasing then decreasing pattern of distance effects across the distances from 1 to 4. The second explanation seems consistent with other research on sequence judgment tasks (e.g., LeFevre & Bisanz, 1986; Lyons &

Beilock, 2009) where counting sequences were processed more quickly and accurately than other sequences.

Lyons and Beilock (2009) investigated both order related judgments for digits in Experiment 2. Because they were interested in individual differences, they measured working memory capacity. They used stimulus sets made up of three single-digit numbers (e.g., 1 2 3; 4 8 5) rather than number pairs. They used three digit sequences because, as suggested by the results of Turconi et al. (2006), participants can make order task judgments on two digits (i.e., a number pair) using relative magnitude without explicitly processing relative order. Thus, decisions on three single-digit numbers would emphasise order and de-emphasize magnitude, although note that they did not explicitly assess whether magnitude judgments were part of the task.

Lyons and Beilock (2009) used three types of stimuli: counting (e.g., 1 2 3), balanced (e.g., 2 4 6), and neutral (e.g., 1 3 4). Thus, the ordered counting stimuli had an inter-digit distance of 1, whereas the ordered neutral stimuli had unequal inter-digit distances within each set. This difference in stimulus construction resulted in a greater number of errors and increased response latencies for the neutral compared to the counting (sequential) stimuli. Participants responded much more quickly to ordered counting stimuli than to the unordered version (e.g., 1 2 3 vs. 2 1 3) whereas differences between ordered and unordered neutral stimuli were minimal (e.g., 3 5 7 vs. 5 3 7). These results suggested that, as in Turconi et al. (2006), immediately adjacent numbers were processed differently than those with larger differences. Lyons and Beilock also noted that on the balanced stimuli, the difference between ordered and unordered versions was larger for individuals with higher working memory than those with lower working

memory, and that working memory skill overall predicted speed of responding. Lyons and Beilock argued that the order judgment task could be very useful for understanding how numerical knowledge was represented, more generally, citing the correlation with working memory ability as potentially linking ordinal processing skill with mathematical proficiency.

Bourassa (2014) conducted two experiments exploring the role of familiarity rather than ordinality in sequence judgments. Four types of three single-digit stimuli were used (counting, balanced, arithmetic, neutral) with two conditions (ascending, descending). Arithmetic fluency (skill) was measured using the Calculation Fluency Test (Sowinski, Dunbar, & LeFevre, 2015). The high fluency group responded more quickly and made fewer errors than the low fluency group. Response latencies were faster for the ascending than the descending sequences. Participants responded fastest to ordered counting sequences {1 2 3}. Bourassa explained that familiar sequences could be recalled directly from long-term memory whereas sequential comparison of the digits is used to determine order on unfamiliar sequences. Consistent with the familiarity hypothesis, participants were slowest to respond to unordered counting sequences. This finding was interpreted as an interference effect. The numbers in an unordered counting sequence {2 3 1} match the components of the familiar sequence {1 2 3}. On these trials, participants presumably had to inhibit the incorrect “Yes, ordered” response, hence slowing the correct “no, unordered” response. Accordingly, Bourassa concluded that rather than ordinality, the familiarity of a number sequence was the chief factor in recognizing the sequence.

Lyons and Beilock (2011) used the same order judgment task as Lyons and Beilock (2009) and Bourassa (2014) to examine the relationship between non-symbolic number discrimination ability and arithmetic skill. They concluded that number ordering ability fully mediated the relationship between approximate number system ability and mental arithmetic ability. Furthermore, the correlation between performance in the order judgment task (combined speed and accuracy) and performance on the arithmetic task was very high, $r(52) = -.703$, in support of the conclusion that arithmetic and order processing were highly related.

In summary, the research on order judgments suggests that sequence processing is an important aspect of number knowledge and is related to arithmetic ability (Lyons & Beilock, 2009, 2011; see also Lyons, Price, Vaessen, Blomert, & Ansari, 2014). Turconi et al.'s (2006) research provides evidence that different processes are used when making decisions about consecutive number pairs versus non-consecutive pairs. Thus, the order judgment task seemed like a reasonable one in which to explore the relation between SNARC effects and order. Turconi et al.'s work suggested that using only two digits in order judgments might not result in clear distinctions between order and magnitude processes, however, whereas three-digit sequences seemed more likely to elicit consistent order judgments. However, Bourassa (2014) showed that participants do use serial search processes to establish order. Thus, in this thesis I used order judgments on three-digit sequences, but I also manipulated the selection of sequences in an attempt to control the processes that participants used across the stimuli.

Contextual Factors that May Influence SNARC Effects

Computational skills

Several researchers have studied the relations between arithmetic skills and SNARC effects. Clark and Campbell (1991) stated that the parity of a number is determined through mental arithmetic. They went on to predict if an individual is slow at calculating, then they must also be slow at judging parity. To test the prediction, Dehaene et al. (1993) in their first experiment categorized participants according to their area of study and included it as a proxy for math skill in their analyses. Ten participants were enrolled in literary studies and 10 participants were enrolled in mathematics, physics, or biology. They performed a standard parity task (i.e., judging whether a single-digit number from the number set 0 – 9 was odd or even). A SNARC effect was found for both the literary and the scientific groups. The slope of the SNARC curve for the Literary group was -10.1 whereas the slope of the curve for the Scientific group was -4.1, leading Dehaene et al. to suggest that participants taking science-related had better number knowledge than those enrolled in literary studies. However, this difference was never tested statistically and, there was no direct testing of number knowledge or calculation skills. It is also not entirely clear why individuals with good mathematical knowledge would show less of a SNARC effect.

In the first of two experiments investigating the inclusion of negative numbers on the number line, Fisher and Rottmann (2005) included experience with numbers as a factor. Twenty participants took part in the parity task of Experiment 1: 10 were studying math related courses and 10 were studying psychology. The same SNARC effect was

found for the two groups. No direct measure of math skill was conducted and the mean response times for both groups were not significantly different.

Link, Nuerk, and Moeller (2014) presented children in grade four with two physical number lines. The bounded number line had two identified positions, the beginning (0), and the end points (1000). The unbounded number line only identified the beginning position (0). The task was to mark the location of a presented number on the line. Measures of the students' calculation fluency were also gathered. To mark a location of a number it was thought that the unbounded line would require the child to access a spatial representation and use that representation to mark the number. This task would be a better indicator of a mental number- space association. Though a positive association between arithmetic skill and the task of marking the number on the appropriate location of a bounded number line was found, there was no such association for the unbounded number line. The spatial-numerical association was limited, even in children with a relatively familiar task. The simple exclusion of an end point on a number line negated the influence of arithmetic skill.

Another investigation into the influence of math skill on the SNARC effect was conducted by Cipora and Nuerk (2013). They used a much larger sample size than Dehaene et al. (1993) or Fischer and Rottmann (2005), testing 71 individuals instead of 20. Participants were once again drawn from math-related and non-related courses. However, in this study the hypothesized skill difference was verified with an arithmetic test. Participants judged whether equations (e.g., $6 + 4 = 11$) were true or false as index of arithmetic fluency. Consistent with the two previous studies, Cipora and Nuerk (2013) did not find a significant correlation between math skills and the SNARC effect.

In a review of spatial-numerical associations, Cipora, Patro, and Nuerk (2015) concluded that associations between arithmetic skill and spatial-numerical associations were not found in all tasks and contexts. The association between arithmetic skill and spatial associations was found in children for numerical intervals, suggesting that the relationship between the SNARC effect and arithmetic skill was only found when the participants with very low and very high arithmetic skill were tested (Cipora et al., 2015).

In contrast to the studies discussed so far, Hoffman, Mussolin, Martin, and Schiltz (2014) reported a significant relationship between mathematical skill and the SNARC effect. The researchers categorized three groups of participants based on math experience and skill: those who self-declared as having difficulties with mathematics, those taking math-related courses, and those taking non-math-related courses. These groupings were subsequently verified through: A) a paper and pencil test with 80 math-related questions, and B) a computerized, timed mathematics-verification test that included addition, subtraction, and multiplication. All three groups demonstrated a SNARC effect when performing the traditional parity judgment task but the mean slope of the SNARC effect was steepest for the group with the lowest arithmetic scores (self-declared math difficulties), and shallowest for those with high scores. This pattern is somewhat puzzling because it seems more likely that participants with stronger number skills would have more knowledge about numbers, including spatial relatedness. Hence, those in math-heavy courses should be more susceptible to SNARC effects, rather than the reverse. One possible explanation is that the math group with the highest math scores directly recalled parity facts from memory and were also able to inhibit the irrelevant number information.

In discussing the contrasting findings, Hoffman et al. (2014) noted that Cipora and Nuerk (2013) had compared unbalanced groups (those taking math or non-math related courses), both in terms of overall number between each group (1:3 odds ratio), and by gender within each group. There were more males than females in the math courses, and the reverse was the case for the non-math courses. Hoffman et al. (2014) argued that the gender imbalance within and between groups might have masked any math-skill differences. They cited research (Bull et al., 2013) showing a larger space-number effect for males than females. The gender imbalance between the math-related groups may have diminished any difference in average slopes between the two groups.

In summary, across four studies where the SNARC effect was compared for groups of individuals varying in experience with math (measured in various ways) only one showed significant effects of experience. Two others showed the same trend however, where the SNARC effect was less for those with more experience. This pattern will be explored in the present research.

Mathematics education

Computational skill is often a result of mathematical education. The quality, focus, and amount of time spent on learning and practicing arithmetic may vary between countries. As such, math skill could also depend on where the participant received their mathematical education. Sowinski et al. (2015) noted that participants who had learned their arithmetic in China correctly solved significantly more calculation problems on a timed calculation test than participants educated in Canada, or elsewhere. Similarly, Campbell and Xue (2001) and LeFevre and Liu (1997) found cultural differences in mathematical skill with Chinese students outperforming non-Asian Canadians.

Number Language

Another important contextual factor in the SNARC effect involves the habitual direction of reading and writing of the participants. In Dehaene et al. (1993), participants were university students in France. Some of the students, though, were born in Iran and had arrived in France between 1.5 years and 12 years previous to the experiments. Whereas the French write from left to right, Iranians write from right to left, although both groups write numbers from left to right. Dehaene et al.'s (1993) Experiment 7 was run in order to determine if the SNARC effect might be modulated by the direction of writing. Twenty Iranian students did a parity judgment task. Although the SNARC effect was nearly significant when the results of all 20 participants were pooled across the two number types, the result was reported as nonlinear, and further analyses indicated that the SNARC effect was modulated by the length of time participants had been in France, or how long ago they had started learning French. These effects suggested that first or dominant language direction may influence the strength or direction of the SNARC effect.

Language differences have also been found when judging which of two remembered, but not seen, animals was either smaller or larger (Shaki et al., 2012). Two different cultural groups made judgments about the comparative size of remembered, but not seen, animals. Twelve participants were left-to-right readers from Canada. Twelve Israel and Palestinian participants from Israel were right-to-left readers. Participants judged which of the two animal names represented larger animals on half the trials, and which represented smaller animals on the other half of the trials by pressing the left button on a keyboard on half the trials, and the right button on the other half of the trials.

When the Canadian participants chose which was smaller, a significant SNARC effect was found. When they chose which was larger, a marginal reversed SNARC was found. In contrast, the direction of these two slopes was reversed for the Israelis and Palestinian participants, indicating that the direction of the SNARC effects reversed when the reading direction reversed.

Shaki et al. (2012) also asked participants (a new group of 12 in Canada, but the same participants in Israel) to make comparative judgments of small and large number pairs. The instructions were the same as in the previously described experiments with the names of animals. There was a SNARC effect for the judgments by the left-to-right readers for both the “choose the larger” and “choose the smaller”. A cultural, or direction of reading effect was found both for comparative judgments of numbers and animal names.

Zebian (2005) explored the influence of writing direction on the SNARC effect by recruiting Arabic- and English-speaking monoliterates and biliterates, as well as Arabic illiterates. The 19 monoliterate Arabic participants, who wrote and read from right to left were tested outside the laboratory, often in their own homes. Not only were the testing conditions different, but also experiences and culture necessitated adaptations to the testing process. These included training periods, and unscheduled participant-initiated breaks in the testing, which the experimenter noted. The task was a same/different task using the four numbers 1, 2, 8, and 9. Two numbers were displayed on the laptop screen, one number to the left and the other to the right of center. Ten number pairs were constructed (e.g., 1-1, 1-2, 1-8, 8-9) resulting in same number pairs, small number pairs, mixed magnitude pairs, and large number pairs. Each pair was presented in each left-right

order. The presented stimuli were responded to orally, by saying ‘yes’ if the two numbers were the same, and ‘no’ if the two numbers were different. Zebian found a reverse SNARC effect for the Arabic monoliterates (i.e., faster different responses to pairs with the smaller number on the right). For the biliterates (Arabic-French, and Arabic-English) there was a non-significant trend towards such a reverse SNARC. Zebian (2005) interpreted this as a weakening of the reverse SNARC by the second language.

Curiously, the SNARC effect in Zebian’s study did not reach significance for the group made up of 19 English monoliterates. A follow-up experiment was conducted for the English monoliterates to address the possibility that the effect was dependent on use of hands, and that an oral response had removed the effect. Twenty English students performed the same task, assigning ‘yes’ and ‘no’ to left/right and right/left keys on the laptop. The proportion of each type of response was not balanced (one third ‘yes’ and two-thirds ‘no’). Again, the SNARC effect did not reach significance ($p = .10$). However, when the experiment was modified again so as to have an equal number of spoken ‘yes’ and ‘no’ responses, of the four types of number pairs presented (now made up of a combination of number and number word pairs, in two orders), only the effect for the number word (to the left) and number (to the right) pairs reached significance. SNARC effects were found for oral responses for Arabic participants. SNARC effects for English monoliterates were only found when responses were made manually. More importantly, Zebian (2005) concluded that the direction of her SNARC effect was a cultural artefact of the habitual direction of reading and writing.

Fischer, Shaki, and Cruise (2009) tested language direction effects with Israeli participants. Nineteen university students took part in their experiment, all of which had

been born in the USSR and immigrated to Israel. Their first language was Russian and they had been speaking Hebrew for at least six years. Participants completed parity judgment experiments with number word stimuli in both languages. A number word in either Cyrillic (which is typically read from left to right) or Hebrew (which is typically read from right to left) was displayed in the center of the monitor. On the trial following each number word, an Arabic digit (which is read from left to right in both languages) was presented in the center of the monitor to which the participant was to respond in the same manner as the trials displaying the number words. This design allowed for analyses to be conducted for digits followed by words in both languages. Subsequently, the difference in response times for the digits were analyzed according to the language of the preceding word. A SNARC effect was found for Russian number words and also for numbers that were preceded by a Russian word. In contrast, no SNARC effect was found for Hebrew number words or for numbers that followed a Hebrew word. Fischer et al. expected this result for Hebrew numbers in the context of preceding Hebrew words because the reading directions for each type of stimuli are in conflict. In a second experiment, participants heard the number words, removing the previously reported possible conflict due to reading-direction. The SNARC effect was now significant for both languages, though the slope for the Russian number words was steeper than the slope for the Hebrew number words. Fischer et al. interpreted the overall results as evidence that the SNARC effect was responsive to recent spatially-related experiences. Individuals seem to be capable of maintaining two language-related mental maps and are able to switch back and forth between them.

Hung, Hung, Tzeng, and Wu (2008) conducted a dual-language parity judgment task with Taiwanese university students to determine if the influence of reading direction on the SNARC effect extended to the vertical plane. All participants spoke Chinese. They were also familiar with English writing by the time they were 13 years of age. Arabic digits and two types of Chinese number words were used, simple form (less formal, more often used) and the complex form (more formal, less used). Reading directions for the two types of stimuli differed. Arabic digits are read left to right but Chinese characters (the number words) are usually read top-to-bottom. In Experiment 1, all three number types were presented centrally. The participants had to make the standard parity judgment, with responses being made horizontally on a response box, with response-directions (odd-left and even-right) switched halfway through the set of trials. The number magnitude by side-of-response interaction was significant for the Arabic digits, but not for the simple or complex Chinese number words. Experiment 2 followed the same procedure except that the responses were now made using vertically aligned keys on a keyboard. A vertical SNARC effect was found for responses made to the simple Chinese words, but not for either the complex Chinese words or the Arabic digits. Hung et al. (2008) concluded that people do create an analog mental representation, that is, they map familiar numbers onto the most familiar spatial pattern. For Taiwanese university students, Arabic digits are mapped onto a number line that flows from left to right. For simple Chinese number characters, however, the number line falls from top to bottom.

Ito and Hatta (2004) also wondered if a SNARC effect could be found for Japanese participants, and if so, would it follow the vertical top to bottom direction of the written language? Working with Japanese university students in Japan, the classic parity

judgment task was used in their Experiments 2A and 2B. In Experiment 2A the left hand was assigned to the bottom button, and the right hand to the top button of the response panel. This hand-to-key assignment was reversed in Experiment 2B. The assignment of odd and even to the response buttons was counterbalanced between two blocks of number presentations. In both experiments, the SNARC effect was significant. However, Ito and Hatta found that large numbers were associated with the top of the analog mental representation and small numbers with the bottom of the analog mental representation (or vertical number line), the opposite of the direction of reading in Japanese, which is top to bottom. Ito and Hatta proposed that this result could be explained by a natural directionality that arises from life's experiences, such as filling glasses with liquid or stacking blocks (Tversky, Kugelmass, & Winter, 1991).

Taken together, the findings from comparisons across language and cultural groups (i.e., Fischer et al., 2009; Hung et al., 2008; Ito & Hatta, 2004; Zebian, 2005), demonstrate that the SNARC effect is sensitive to cultural influences such as the direction of writing, but also to well-remembered natural phenomena, such as water rising in a glass in the case of vertical contexts, and to within-experiment contextual effects, such as the direction of reading for a word on a previous trial. Accordingly, in the three experiments I present in this thesis, I documented the language experience of each participant.

Analyzing SNARC

There is some variability in how the SNARC effect is analyzed across different studies. The initial method, used by Dehaene et al. (1993) was to conduct a repeated-measures ANOVA and determine if the interaction of number magnitude by side of

response was significant. However, Fias, Brysbaert, Geypens, and d'Ydewalle (1996) suggested that a two-stage analysis incorporating a method for testing slopes in within-subject designs as described by Lorch and Myers (1990) "... capture[s] the essence of the SNARC effect more clearly" (p. 99). In this method, the data for each participant is analyzed by regressing difference scores (right hand RT – left hand RT) on number magnitude (e.g., 1 through 9) to calculate a slope (i.e., a SNARC effect). The slopes are then tested against zero (in a single-sample t-test) to determine significance. Fias et al. listed a number of advantages to the Lorch and Myers method. First, the t-test evaluates a main effect, as opposed to an interaction, making it more easily comparable across different samples. Second, the factor of interest (e.g., magnitude from 1 to 9) is treated as a continuous variable and does not have to be dichotomized. Third, the magnitude of the slopes under different factors or conditions can be directly compared. Fourth, the slopes (unstandardized coefficients) can be compared across between-subject factors such as language, gender, or skill. Fifth, because the slopes are generated for each participant, any distortion due to group averaging is reduced. And sixth, the method "does not require an orthogonal combination of the factors in the design" (p. 100). The slope coefficients calculated using this method of individual regressions are used in further analyses. As explained above, they can be compared across groups who speak different languages or correlated with a continuous measure such as calculation fluency.

As an example of how the two-step regression approach can be applied to data in the field of numerical cognition, LeFevre and Liu (1997) used a similar approach to analyzing solution latencies to multiplication problems. They used regression on individual participants to calculate slope values for problem size (i.e., product), problem

types (i.e., ties versus non-ties; problems with 5 as an operand) and then analyzed the coefficients in ANOVA designs. More recent approaches involve the use of multi-level models to capture both between- and within-group variance simultaneously, though this technique has not yet been used to analyse possible SNARC effects. In the present research, I first present traditional analyses of response times and percentage of errors using ANOVA analyses because the order judgment task is novel and thus patterns of relations among the variables was of interest. For those analyses, SNARC effects would be expected as interactions between number size (e.g., 1 2 3 through to 7 8 9) and hand-of-response. Second, I used Lorch and Myers' approach to calculate unstandardized coefficients as indices of potential SNARC (or SNARC-like) effects. When appropriate, I calculated the coefficients separately in different experimental conditions and compared across conditions or groups (e.g., language groups). Finally, where appropriate, I used correlational analyses with measures of skill to explore individual differences in the SNARC effects.

Power Calculations

Wood, Nuerk, and Willmes (2006) combined the results from seven experiments involving research into the SNARC effect to arrive at an estimate of the unstandardized Beta coefficient (-4.93 ms per unit increase), and the standard deviation (8.27). Based on these estimates of effect size and variance, and using a more conservative two-tailed test ($\alpha = .05$), 24 participants would be needed to achieve a power of 83% in a typical SNARC experiment. For 90% power, 30 participants would be needed. Thus, in Experiments 1 and 2 I recruited 34 participants. Note that this sample size is about twice as large as in many other SNARC studies. In Experiment 3, my goal was to explore

relations between other individual difference variables and the SNARC effect, and so I increased my sample size to 73 to maximize the value of that approach.

Summary of Hypotheses

To summarize, Dehaene et al. (1993) had concluded that magnitude, and not order, drove the SNARC effect for numerical stimuli. Other researchers increased the number of participants, improved the experimental tasks, and demonstrated that a SNARC, or at least a SNARC-like, effect was found for tasks that involved a variety of ordered non-numeric sequences for which the items were presented in single-stimuli judgment and comparison experiments (Gevers et al., 2003; Gevers et al., 2004; Prado et al., 2008; Previtali et al., 2010; van Opstal et al., 2009). Importantly, Gevers et al. (2003) stated that a possible problem with the original order experiment (i.e., Experiment 4 of Dehaene et al., 1993) was that order had not been relevant to the task. Thus, SNARC effects may occur whenever order information is activated as part of the task demands or, in the case of numbers, when order is an inherent feature of the domain. Accordingly, examining numerical tasks that explicitly require participants to activate order should contribute to our understanding of SNARC effects.

Just as Dehaene et al. (1993) used the lack of a SNARC effect for letters to conclude that magnitude and not order was complicit in SNARC, the same logic was used in the experiments described in this thesis. Participants were asked to judge if single-digit number triads (e.g., 1 2 3 or 2 1 3) were in order or not. If magnitude is the numerical factor that drives the SNARC effect then SNARC patterns should occur for both ordered and unordered sequences because magnitude information is equally available from both sequences. In contrast, if activation of order information is the source

of SNARC, then I expected to find SNARC for the ordered but not the non-ordered sequences. Further, since order is equally implicated in ascending and descending sequences (e.g., 2 3 4 and 4 3 2), then both ascending and descending ordered sequences are expected to show SNARC effects. These hypotheses were tested in the three experiments, as listed in Table 2.1.

Experiment 1 showed some results consistent with these hypotheses. Asian-educated individuals showed a significant SNARC effect for ascending ordered triads. The English group showed a SNARC effect for descending ordered triads, and a reverse SNARC for ascending ordered triads. Hence, in Experiment 2 the stimuli were revised to determine if they were the reason for a lack of a significant SNARC effect for unordered triads. As well, a calculation test was added as math skill and not just country of origin has been implicated in SNARC. The results of Experiment 2 confirmed those of Experiment 1. A SNARC effect was found for the high math skill group, but not the low math skill group. Again, SNARC was found for ascending ordered triads, but not descending sequences. In Experiment 3, the order task was limited to the ascending task but some participants implicitly learned global ordering of the stimulus sets. All participants within Experiment 3 also performed a standard parity judgment task in order to compare SNARC effects across tasks. SNARC was found for individuals with high calculation skill, and for non-ordered sequences for those individuals who had undergone training on the global ordering of the sequences. However, the correlation between SNARC for parity judgments and SNARC for sequence judgment was not significant. Table 2.1 lists the hypotheses for each of the three experiments.

In summary, the three experiments described in this thesis examined the question of whether the hand by side-of-response bias known as the SNARC effect that has been shown for numerical tasks is related to the activation of order information, rather than of magnitude as has been assumed. It is clear that SNARC-like patterns are shown in other tasks that involve order judgments, but for numerical tasks, it has been assumed that magnitude has been activated explicitly (e.g., in magnitude comparison tasks) or implicitly (in parity tasks) and causes the observed SNARC effects. Parity tasks, however, may also activate order information so the presence of a SNARC effect does not prove that magnitude is the cause. By using a numerical task that explicitly activates order information and evaluating whether SNARC-like effects are observed in relation to number magnitude independently of order, I hope to determine whether order plays a role in the so-called SNARC effect found in numerical tasks.

Table 2.1 Summary of thesis experiments and hypotheses

Experiment	Hypotheses
1	<ul style="list-style-type: none"> • A SNARC effect will be found for judgments of numerical order of ordered sequences.
	<ul style="list-style-type: none"> • A SNARC effect should be found for both ascending and descending ordered sequences.
	<ul style="list-style-type: none"> • There will be no SNARC effect for non-ordered sequences.
2	<ul style="list-style-type: none"> • Was the lack of a SNARC effect in Experiment 1 for unordered sequences and for descending sequences due to the differences in how the triads were constructed?
	<ul style="list-style-type: none"> • Was the finding of the SNARC effect in Experiment 1 for Asian (vs. non-Asian) students due to language effects or to differences in math skill?
3	<ul style="list-style-type: none"> • Can training the global order of the unordered triads result in a SNARC effect for unordered triads?
	<ul style="list-style-type: none"> • Are the SNARC effects for parity and sequences the same?

CHAPTER 3 – EXPERIMENT 1

In an attempt to separate the roles of magnitude and order when using numbers, participants in this experiment were asked to judge whether three single-digits (i.e., Triads) were in order (e.g., 1 2 3) or not (e.g., 2 1 3). Triads have been used before to study order processes (Lyons & Beilock, 2009) but have not been used in SNARC-related experiments. If the SNARC effect is driven by magnitude, it should occur on both the ordered and unordered Triads. However, if order, and not magnitude, plays the more important role in the SNARC, then the ordered and unordered stimulus sets may show different effects. Specifically, my first hypothesis was that there would be a SNARC effect for ordered, but not for unordered Triads.

In this experiment, I also manipulated sequence direction, that is, whether triads were increasing (e.g., 1 2 3) or decreasing (e.g., 3 2 1). Sequence direction was manipulated across blocks but within participants. In relation to SNARC, if magnitude is the critical feature, then there should be a SNARC effect for both the increasing and decreasing stimuli. Bourassa (2014) found, however, that order effects were smaller on decreasing than on increasing sequences and thus, if order drives the SNARC effect, then I hypothesized that it may be smaller or absent on decreasing sequences.

An alternative prediction is that increasing stimuli will induce a rightward bias in participants whereas decreasing stimuli will induce a leftward bias (Bächtold, Baumüller, & Brugger, 1998). This prediction is also consistent with research on the operational momentum effect, in which addition problems induce a rightward bias, presumably because they involved moving rightwards along the number line whereas subtraction problems induce a leftward bias (Fischer & Shaki, 2014; Knops, Zitzmann, & McCrink,

2013). Thus, my third prediction was that sequence direction and hand-of-response would interact if sequence processing activated directional processes that are related to SNARC effects (Masson & Pesenti, 2014).

A final consideration in the current research was the finding that SNARC effects are related to typical direction of writing (e.g., Zebian, 2005). A standard SNARC effect (negative slope) is found for participants who read from left-to-right, whereas a reverse SNARC (positive slope) is found for participants who read from right-to-left. Thus, to the extent that the SNARC effect reflects directional biases related to language, a reverse SNARC may be found for Arabic speakers (Shaki, Fischer, & Petrusic, 2009; Zebian, 2005). As such, languages of participant were recorded and used as a between-subjects factor in the analyses.

Method

Participants

Thirty-four students from Carleton University took part in the experiment in exchange for course credit. Data from 3 students was removed due to high error rates and failure to follow instructions. Of the 31 remaining participants, 19 were female, and 28 were right-handed, with a mean age of 21.7 years (*SD* 2.7). The participants' first languages were English (11), Arabic (7), and Asian (13; one of Chinese, Japanese or Korean). Many of the participants were multi-lingual. Cumulatively, the participants were fluent or conversant in 12 languages (Table 3.1).

Table 3.1. Experiment 1: Summary information of first languages of participants, the direction of reading in the first language, and additional language fluencies.

First Language	<i>n</i>	Reading Direction	Additional Languages Spoken
English	11	left to right	French, Hindi, Japanese, Korean, Romanian, Russian, Tamil
Asian (Chinese, Japanese, Korean)	13	left to right	English, French
Arabic	7	right to left	English, French, Pashto, Turkish

Equipment

The experiment was conducted using an IBM Pentium computer, an NEC MultiSync LCD 2060NX monitor, and a modified keyboard. All the keys, except for A, L, and the Space bar had been removed. Stimuli and instructions were presented in the center of the monitor using black type on a white background. The stimuli were controlled and the results collected using Superlab version 4.5 software. Participants sat at an approximate distance of 56 cm from the monitor. The stimuli, in 20 point Tahoma font, were made up of three single-digit numbers with three spaces between each stimulus set taking up a width of 3.5 cm. The visual angle of the stimuli subtended 3.6°.

Materials

The stimuli, or Triads, as seen in Table 3.2, were constructed from the numbers {1 2 3 4 6 7 8 9} omitting the center number 5. The goal was to control the overall magnitude of triads such that it increased across sets. Triad 1 was made up of the first three numbers in the set (i.e., 1 2 3) and Triad 6 was made up of the last three numbers in the set (i.e., 7 8 9). Triads in set 2 were made up of three of the first four numbers (i.e., 1

2 3 4), excluding Triad 1. Triads in set 3 were made up of two numbers below 5, and one number above 5. Triads in set 4 included one number below 5, and two numbers above 5. Triads in set 5 included three of the last four numbers (i.e., 6 7 8 9), excluding Triad 6. Ascending and descending triads were the reverse of each other. The unordered stimuli were used in both ascending and descending task blocks. Ordered triads for ascending and descending tasks were repeated three times per block.

Table 3.2. Experiment 1: Single-digit number sets (Triads).

Sequence Type	Task	Triad 1	Triad 2	Triad 3	Triad 4	Triad 5	Triad 6
ordered	ascending	1 2 3	2 3 4	1 2 6	1 6 8	6 7 8	7 8 9
			1 2 4	2 4 7	4 7 9	6 7 9	
			1 3 4	3 6 7	4 6 7	6 8 9	
	descending	3 2 1	4 3 2	6 2 1	8 6 1	8 7 6	9 8 7
			4 2 1	7 4 2	9 7 4	9 7 6	
			4 3 1	7 6 3	7 6 4	9 8 6	
unordered		1 3 2	1 4 2	1 8 4	2 9 7	6 8 7	7 9 8
		2 3 1	1 4 3	2 7 1	3 6 7	6 9 7	8 9 7
		3 1 2	2 4 1	2 9 4	3 7 6	6 9 8	9 7 8
			2 4 3	3 6 1	3 9 8	7 8 6	
			3 4 1	3 8 1	4 8 7	7 9 6	
			3 4 2	4 7 2	6 9 1	8 6 7	
			4 1 2			8 9 6	
			4 1 3			9 6 7	
			4 2 3			9 6 8	

Note. Triads 1 and 6 for ordered sequences (ascending and descending) were repeated three times.

Procedure

Half of the participants began with the ascending task and half began with the descending task. Within each task, initial response key assignment was also counterbalanced, with half beginning with the letter A (left side) being assigned to “yes ordered”, and the letter L (right side) assigned to “no unordered”. In the second half of the experiment, response key assigned was reversed. Blocks were made up of 36 ordered and 35 unordered trials. Each block was repeated three times for each side of response within each of the two tasks. Trials were presented randomly within each block. Thus, participants saw four blocks in a 2 (side of “ordered” response) by 2 (ascending, descending) design, hence, 71 trials x 4 conditions x 3 repetitions = 852 trials in total. The experiment took approximately 45 minutes.

Participants were asked to judge whether the three single-digit numbers were in ascending (or descending) order. Participants performed six practice trials with feedback before the ascending task, and six practice trials with feedback before the descending task. If the response on a trial was incorrect, a notice appeared repeating the instructions, and the trial was repeated. The practice continued until all of the practice trials had been answered correctly once. Participants were asked to be as fast and as accurate as possible. Participants were asked to take breaks, once between the two tasks, and twice between the response key reversals within each task. This procedure resulted in three self-timed breaks.

During the experiment an orienting cue made up of three asterisks (* * *) appeared on the monitor for 1000 ms, then was replaced by a three-digit sequence in the same location. The trial ended when the participant pressed a response key (A or L). If a

participant responded in less than 200 ms, a message appeared for two seconds asking the participant to please remember to be as accurate and as fast as possible.

Analysis

Data Trimming

Mean correct responses averaged 94.7% across all trials. Participants' mean percentage errors ranged from 0.7% to 17.2%, with a median percentage error of 3.8%. After removing errors and responses that were less than 200 ms or greater than or equal to three times the individual mean response times, 92.8% of the responses were available for analysis.

Speed-Accuracy Trade-off

The correlation between percentage errors and mean correct response times was not significant across individuals, $r(29) = .23$, $p = .217$. As such, there was no speed-accuracy trade-off.

Error Analysis

The percentage errors were analyzed in a 2(Task: ascending, descending) by 2(Stimulus type: ordered, unordered) by 6(Triad 1 – 6) by 2(Side of 'Yes' response: right, left) by 3(First Language: English, Arabic, Asian) mixed ANOVA, with First Language as a between-subjects factor. The Greenhouse-Geisser epsilon adjustment of degrees of freedom was used if the assumption of homogeneity of variance was violated.

There was a main effect of Language. English first-language speakers (6.7%, $SE = 0.98$) made more errors than Asian first-language speakers (3.2%, $SE = 0.90$). The Arabic first-language speakers' mean errors did not differ from either group (5.5%, $SE =$

1.2). The main effect of Task was marginal. More errors were made on the ascending task (5.6%, $SE = 0.71$) than the descending task (4.7%, $SE = 0.60$).

Table 3.3. Experiment 1: Source table for repeated measures ANOVA on percentage errors

Source	<i>df</i> 1	<i>df</i> 2	<i>F</i>	<i>p</i>	η_p^2
Language (English, Asian, Arabic)	2	28	3.54*	.043	.202
Task (ascending, descending)	1	28	3.28+	.081	.105
Task x Language	2	28	0.74	.486	.050
Stimulus Type (ordered, unordered)	1	28	13.60**	.001	.327
Stimulus Type x Language	2	28	0.58	.567	.040
Triad (1 through 6)	3.0	84.2	21.90***	<.001	.439
Triad x Language	6.0	84.2	1.62	.153	.103
Side	1	28	0.10	.751	.004
Side x Language	2	28	0.87	.428	.059
Task x Stimulus Type	1	28	4.46*	.044	.137
Task x Stimulus Type x Language	2	28	1.51	.238	.097
Task x Triad	2.6	72.8	1.15	.332	.039
Task x Triad x Language	5.2	72.8	0.71	.622	.048
Stimulus Type x Triad	3.2	90.1	7.97***	<.001	.221
Stimulus Type x Triad x Language	6.4	90.1	3.30**	.005	.191
Task x Stimulus Type x Triad	3.0	83.9	1.01	.393	.035
Task x Stimulus Type x Triad x Language	6.0	83.9	0.34	.913	.024
Task x Side	1	28	0.51	.482	.018
Task x Side x Language	2	28	0.41	.667	.029
Stimulus Type x Side	1	28	<0.01	.930	<.001
Stimulus Type x Side x Language	2	28	0.43	.656	.030
Task x Stimulus Type x Side	1	28	1.77	.194	.060
Task x Stimulus Type x Side x Language	2	28	1.08	.353	.072
Triad x Side	3.5	98.9	1.75	.153	.059
Triad x Side x Language	7.1	98.9	1.52	.169	.098
Task x Triad x Side	4.1	115.9	0.76	.556	.026
Task x Triad x Side x Language	8.3	115.9	2.30*	.024	.141
Stimulus Type x Triad x Side	3.6	100.6	1.10	.357	.038
Stimulus Type x Triad x Side x Language	7.3	100.6	0.84	.560	.057
Task x Stimulus Type x Triad x Side	4.3	121.2	0.65	.640	.023
Task x Stimulus x Triad x Side x Language	8.7	121.1	1.50	.158	.097

+ $p = .081$, * $p < .05$, ** $p < .01$, *** $p < .001$

There was a main effect of Stimulus Type. More errors were made for unordered (6.3%, $SE = 0.88$) than ordered (3.9%, $SE = 0.41$) sequences. There was a main effect of

Triad. The number of errors generally increased with Triad. The linear trend was significant, $F(1, 28) = 45.86, p < .001, \eta_p^2 = .621$. The interaction between Task and Stimulus type was significant (Fig. 3.1). Participants made more errors on unordered trials in the ascending task than in the descending task, whereas there was no difference across task types for ordered trials.

Figure 3.1. Experiment 1: Interaction plot of Task by Stimulus type on percentage errors. Error bars are 95% CIs.

The significant interaction between Stimulus Type and Triad was qualified by the three-way interaction of these factors with Language (Fig. 3.2). The English first-language participants made significantly more errors on Triads 5 and 6 for the unordered sequences than the Arabic first-language participants, who made more errors than the Asian group.

Figure 3.2. Experiment 1: Interaction plot of Stimulus type by Triad by Language on percentage errors. Error bars are 95% CIs.

Figure 3.3. Experiment 1: Interaction plot of Task by Triad by Language on (right – left hand) percentage errors. Positive values indicate more errors with the right hand than the left. Error bars are 95% CIs.

There was a significant 4-way interaction between Task, Triad, Side of response, and Language (Fig. 3.3). The interaction was plotted using the difference in percentage errors made between the right and the left hand responses. The English group made more errors with their right than their left on Triad 5 of the descending task, but made more errors with their left than their right on the same triad of the ascending task. The Asian group made more errors with their left than their right hands on Triad 4 of the ascending task. The Arabic group made more errors with their right hands than their left on Triads 1 and 2 of the descending task, but reversed this for the ascending task.

The Arabic group also made more errors with their right hands than their left on Triad 4 of the ascending task. The Arabic group made more errors with their right hand than their left, whereas the Asian group made more errors with their left hand than their right on Triad 4 for the ascending task. This reversal of hand errors between language groups was also seen between the English and Arabic groups for Triad 1, with the descending task. The Arabic group made more errors with their right than their left, and the English group made more errors with their left than their right hands on Trial 1 for the descending task. This reversal of percentage errors on task and hand may indicate number-related language processing differences.

Error Summary

Overall, most participants made few errors (median = 3.8%). Participants found the unordered sequences in the ascending task to be relatively difficult. The English participants made more errors than the Asian participants, particularly on the last two Triads. The difference in errors made between the right and the left hand may have indicated processing differences due to reading direction.

ANOVAs on Mean Response Latencies

Individual differences. In order to select variables to include as between-subject factors, initial analyses were done using gender and first language. A one-way ANOVA was conducted on participant mean latencies with gender as the independent variable. There was no significant difference in response times between male ($n = 12$) and female ($n = 19$) participants, $F(1, 29) = .837, p = .368, \eta^2 = .028$. As such the two genders were collapsed for the purpose of further analyses.

Similarly, a univariate ANOVA with First Language (English, Asian, Arabic) was conducted on mean latencies. There was a significant effect of First Language, $F(2, 28) = 4.01, p = .029, \eta^2 = .223$. The Bonferroni adjustment was used for multiple pair-wise comparisons. The Asian group (945 ms, $SE = 128$) responded more quickly than the Arabic group (1557 ms, $SE = 175$), whereas the English group mean response time was not significantly different from the other two groups (1114 ms, $SE = 140$). Accordingly, language was used as a between-subjects factor in the analysis.

Latencies. Mean correct latencies were analyzed in a 2(Task: ascending, descending) by 2(Stimulus type: ordered, unordered) by 6(Triad: 1 - 6) by 2(Side of 'Yes' response: left, right) by 3(Language: English, Asian, Arabic) ANOVA, with repeated measures on the first four factors. The Greenhouse-Geisser epsilon adjustment of degrees of freedom was used if the assumption of homogeneity of variance was violated. The Sidak adjustment was used for multiple pair-wise comparisons.

Table 3.4. Experiment 1: Source table for the within-subjects ANOVA on mean response time.

Source	<i>df1</i>	<i>df2</i>	<i>F</i>	<i>p</i>	η_p^2
Language (English, Asian, Arabic)	2	28	2.01	.153	.125
Task (ascending, descending)	1	28	2.11	.157	.070
Stimulus Type (ordered, unordered)	1	28	41.38***	<.001	.596
Triad (1 through 6)	1.7	49.0	27.77***	<.001	.498
Side	1	28	0.07	.789	.003
Stimulus Type x Triad	2.5	70.1	13.02***	<.001	.317
Task x Side	1	28	0.06	.808	.002
Task x Language	2	28	0.55	.583	.038
Stimulus Type x Language	2	28	0.81	.453	.055
Triad x Language	3.5	49.0	1.13	.349	.075
Side x Language	2	28	1.04	.366	.069
Task x Stimulus Type	1	28	1.61	.215	.054
Task x Triad	2.8	76.9	1.08	.360	.037
Triad x Side	2.0	56.3	0.59	.558	.021
Stimulus Type x Side	1	28	0.21	.651	.007
Task x Stimulus Type x Language	2	28	2.80	.078	.166
Task x Triad x Language	5.5	76.9	1.68	.142	.107
Stimulus Type x Triad x Language	5.1	70.1	1.47	.210	.095
Task x Stimulus Type x Triad	3.5	98.5	2.03	.105	.068
Task x Stimulus Type x Triad x Language	7.0	98.5	0.71	.663	.048
Task x Side x Language	2	28	0.74	.486	.050
Stimulus Type x Side x Language	2	28	1.36	.273	.088
Task x Stimulus Type x Side	1	28	0.40	.528	.014
Task x Stimulus Type x Side x Language	2	28	0.97	.391	.065
Triad x Side x Language	4.0	56.3	1.47	.224	.095
Task x Triad x Side	2.9	81.7	1.23	.287	.044
Stimulus Type x Triad x Side	3.4	94.6	3.54*	.014	.112
Stimulus Type x Triad x Side x Language	6.8	94.6	1.95+	.073	.122
Task x Triad x Side x Language	5.8	81.7	0.83	.541	.056
Task x Stimulus Type x Triad x Side	2.6	73.0	0.40	.721	.014
Task x Stimulus x Triad x Side x Language	5.2	73.0	0.28	.925	.020

+ $p = .073$, * $p < .05$, *** $p < .001$.

As shown in Table 3.4, there was a main effect of Stimulus type. Participants responded faster to ordered stimuli (1073 ms, $SE = 86$) than to unordered stimuli (1215

ms, $SE = 102$). There was a main effect of Triad and the linear trend for Triad was significant, $F(1, 30) = 38.47$, $p < .001$, $\eta_p^2 = .56$. Response latencies generally increased with number size (Fig. 3.4).

Figure 3.4. Experiment 1: Line plot of Response Times for each Triad. The error bars are 95% CIs.

The Stimulus type by Triad interaction was significant (Fig. 3.5) and the linear trend of the Stimulus type by Triad interaction was significant, $F(1, 30) = 14.65$, $p = .001$, $\eta_p^2 = .33$. Response latencies generally increased with number size for both ordered and unordered stimuli. However, the pattern of increase differed across the stimulus types. More specifically, Triads 1, 2, 5 and 6 showed a response time advantage for the ordered stimuli, as compared to the unordered. Triads 1 and 2 are sequences with consecutive numbers, and Triads 2 and 5 include some sequences with consecutive numbers whereas all of Triads 3 and 4 have unequal distances between the digits. Consistent with other research on order judgments, triads with consecutive numbers were

recognized more quickly and accurately than those that did not have consecutive numbers and showed larger differences between ordered and unordered response times (Bourassa, 2014; Lyons & Beilock, 2009).

Figure 3.5. Experiment 1: Interaction plot of Response Times for Stimulus Type by Triad. The error bars are CIs 95%.

The significant Triad by Side of response interaction was qualified by the three-way interaction of Triad x Side of response x Stimulus type (ordered and unordered) as shown in Figure 3.6. The increase across Triad was larger for the ordered than for the unordered stimuli and the pattern of differences across right versus left side of response also varied. These patterns are somewhat clearer in the four-way interaction between Stimulus type by Triad by Side of response by Language (Appendix iii, Figures 3.8 – 3.13). There appears to be a reversal of hand speed advantage for the ordered and unordered stimuli across language groups and stimulus type. The English-language group have faster right- than left-hand response latencies for the unordered stimuli (Appendix

iii, Figure 3.8). But this pattern reverses and the left hand becomes faster for the ordered stimuli (Appendix iii, Figure 3.9). There was a crossover effect for ordered and unordered stimuli for the Arabic group, with the left hand responses slower than the right for Triad 1, but left- faster than right-hand responses for Triad 6 (Appendix iii, Figure 3.10 and Figure 3.11). There is also a possible crossover effect for the Asian group for the ordered stimuli with right- slower than left-hand for Triad 1, but faster than the reverse for Triad 6 (Appendix iii, Figures 3.12 and 3.13). This latter pattern is evidence of a SNARC effect. To better understand these patterns, in the next section *t*-tests were conducted to determine if any of the differences across side of response were significant.

Figure 3.6. Experiment 1: Interaction plot of Response Times for Triad by Stimulus type (ordered, unordered) by Side of response. The error bars are 95% CIs.

SNARC Analysis

SNARC Analysis

Within Experiment 1, the SNARC effect would be reflected in an interaction between Triad (which indexes magnitude) and Side of Response. The interaction between Triad and Side of response was not significant, however, the four-way interaction of Triad by Side of Response x Sequence Type x Language approached significance, suggesting that the SNARC effect may vary with language groups and sequence type. Research has shown that the SNARC effect is sensitive to the order in which the participants' first language is read, that is left-to-right or right-to-left (Fischer, Shaki, & Cruise, 2009; Hung et al. 2008; Ito & Hatta, 2004; Zebian, 2005). In this experiment there were three language groups, however, both the English and Asian groups read left-to-right in their first language, whereas the Arabic group read from right-to-left in their first language. Zebian (2005) found evidence of a reversed SNARC effect in Arabic monolinguals.

Thus, following the Lorch and Myers (1990) Method 3 procedure, differences (i.e., right-hand minus left-hand latencies) were calculated for each participant. Regressions were then conducted for each participant, with the difference in mean response times as the dependent variable, and Triad as the independent variable, separately for each Task (ascending, descending) and for each order condition. A single-sample *t*-test comparing the mean slope to zero was then conducted on the resultant unstandardized slopes for each of the language groups in each condition. In the literature, a one-tailed test is usually conducted on the slopes because the direction of the slope is expected to be negative. In the present research, however, a two-tailed test was

conducted, because it was not known whether a typical or a reversed SNARC effect would be found. The results of these tests are shown in Table 3.5 and the data for ordered sequences are shown in Figure 3.5.

As shown in Table 3.5, none of the slopes for unordered stimuli were statistically significant. Slopes for ordered stimuli were significantly different than zero for the English-speaking and Asian participants but not for the Arabic participants. There was a significant *reversed* SNARC effect for the English-speaking group on ascending trials (Fig. 3.7A), and 73% of the individual slopes were positive (Table 3.5). In contrast, for the descending ordered trials, the English-speaking group showed the expected negative effect (Fig. 3.7B) and 73% of individual slopes were negative (Table 3.5). The positive slope for the ascending task for English-speaking participants was unexpected, because the SNARC effects in languages that are read from left-to-right are expected to be negative. Further, contrary to the hypothesis that descending tasks would lead to weaker, or no SNARC effect, the magnitude of the descending ordered slope was larger than that of the ordered slope.

In contrast, the Asian-language participants showed consistent results, with negative SNARC slopes for both ascending (Fig. 3.7E) and descending (Fig. 3.7F) ordered trials. For the descending ordered sequences, the effect was not significant. As shown in Figure 3.7, the data are less orderly for the descending than for the ascending trials and the SNARC effect was probably driven by the extreme Triads, suggesting that the variability in sequence composition may have influenced participants' performance. Nevertheless, for both ascending and descending tasks, a high percentage of the individual slopes were negative. These findings suggest that the Asian-speaking group

showed typical SNARC effects on the order judgement task for ordered stimuli, on both ascending and descending sequences.

Table 3.5. Experiment 1: Summary of the SNARC effect for the First-Language groups. Type is Stimulus type (ordered and unordered)

Language	Task	Type	Slope	% neg ^a	<i>t</i>	<i>df</i>	<i>p</i>
English							
	ascending	unordered	-7.2	55	-0.66	10	.525
		ordered	17.4	27	2.75**	10	.020
	descending	unordered	-17.5	45	-1.00	10	.342
		ordered	-21.9	73	-2.77**	10	.020
Asian							
	ascending	unordered	-8.2	62	-0.82	12	.428
		ordered	-20.7	85	-2.92**	12	.013
	descending	unordered	-25.1	54	-1.10	12	.295
		ordered	-15.4	77	-1.73+	12	.109
Arabic							
	ascending	unordered	13.2	57	0.63	6	.553
		ordered	32.4	57	0.57	6	.593
	descending	unordered	3.9	86	0.08	6	.942
		ordered	40.9	57	1.09	6	.319

^a Percentage of slopes that were negative across individuals.

+ $p = .109$, ** $p < .01$ (two-tailed)

Figure 3.7. Experiment 1: Line plots of the SNARC effects for ordered stimuli. The lines are the lines of best fit.

In contrast, although the SNARC slopes for the ascending ordered (Fig. 3.7C) and the descending ordered (Fig. 3.7D) for the Arabic group were both positive, neither was

statistically significant and fewer than half the Arabic first language speakers had a positive slope (three of seven) in spite of the positive mean slopes for ascending and descending sequences (Table 3.5). It is important to keep in mind, however, that the Arabic-speaking group was the smallest of the three groups and so there was less power to detect effects than for the other two groups. Further tests of whether the Arabic group show a reversed pattern of SNARC effects in the sequence judgment task would require using a larger sample size.

Power Analysis

A post hoc power analysis, using MorePower 6.0.1 (Campbell & Thompson, 2012) was conducted on the SNARC effects (Appendix A, Table A.1). Power for the significant SNARC effects (Table 3.5) ranged from .69 up to .76. Using the mean standard deviation (24.2) and η^2 (42.7) from three significant effects (but not the marginal effect) there should be at least 14 people in each cell of the next experiment.

Discussion

The goal of Experiment 1 was to test for the presence of a SNARC effect in an order judgment task in which participants decided whether a number set made up of three single-digit numbers (i.e., a Triad) was ordered (e.g., 1 2 3) or unordered (e.g., 2 3 1). Some evidence for a SNARC effect was found for ordered sequences but none for unordered sequences. If magnitude had been the sole reason for a SNARC effect, then the effect should have been found for both the ordered and the unordered sequences. However, the results suggest that only ordered sequences show consistent side-of-response effects and thus, the results are consistent with the hypothesis that order, and not magnitude, is related to SNARC effects in this task. I also hypothesized that there would

be different patterns for ascending and descending sequences. Although there were no interactions with sequence direction, and SNARC effects were found for both ascending and descending sequences, interactions with language group suggest that sequence direction may be related to spatial-number associations. The variability in performance across language group indicates that individual differences may be important factors in the SNARC effect for ordered sequences.

For the ascending task there was a reverse SNARC effect for the Ordered Triads for the English-language participants, whereas they showed a typical SNARC effect on descending sequences. In contrast, the Asian-language group showed typical (negative) slopes for both ascending and descending sequences. Why did the Asian-language participants (Chinese, Japanese and Koreans) have a regular SNARC effect and the English first-language participants show both reversed and typical SNARC effects (depending on sequence direction)? The cause was not due to reading direction, as both English and many Asian participants read from left to right. In particular, modern Chinese reading uses a left-to-right direction.

One possibility is that the differences across language groups are related to differences in arithmetic skill rather than language. Performance on the order task was strongly related to arithmetic skill in other research (Bourassa, 2014; Lyons & Beilock, 2011) such that individuals who were more skilled at arithmetic were faster and more accurate. Furthermore, LeFevre et al. (2014) found that Asian students had better arithmetic skills than Canadian students. Advantages for Asian-educated individuals have been found in several comparisons of Canadian- and Asian-educated students (Campbell & Xue, 2001; LeFevre & Liu, 1997). It is quite likely, therefore, that the Asian-language

group in the present research was more skilled than either of the other groups and so their more stable performance patterns on the sequence judgment task may reflect higher levels of arithmetic skill.

Why might better arithmetic skills be related to SNARC effects for the sequence judgment task? One possibility is that participants with better number knowledge may use direct access to stored knowledge about sequence patterns whereas those with weaker skills might use serial comparison to determine whether digits were ordered (LeFevre & Bisanz, 1996). Serial comparisons result in slower response times and more variability whereas direct recall of number knowledge for sequences such as {1 2 3} is fast and accurate, and may be more likely to retain spatial associations, such as small numbers being automatically associated with left space. Thus, one explanation of the relation between SNARC effects and group is that they reflect strategic differences in sequence processing by individuals with high- versus- lower levels of arithmetic fluency.

This explanation, that is, different use of strategies on the sequence judgment task producing different patterns of performance, is consistent with some results reported by Turconi et al. (2006). Turconi et al. compared performance for pairs of numbers (e.g., 5 vs. 7) when the task was varied, that is, magnitude comparison (e.g., which is larger?) versus order judgments (e.g., are they in order?). For magnitude comparisons, Turconi et al. observed a standard distance effect that is, decreased response time for magnitude comparisons as the distance between numbers increased. In contrast, they found a reverse distance effect for order judgments that is, faster response times as the distance between numbers decreased. This reversal for order judgements was limited to stimulus pairs made up of consecutive numbers (e.g., 5 6).

Turconi et al. (2006) presented two possible explanations for the reversal of the distance effect with consecutive numbers. First, participants might have used serial search for consecutive numbers (e.g., 3 4) but switched to magnitude comparison for numbers that differed by more than one (e.g., 3 5). Second, they suggested that magnitude comparison is actually the default strategy for comparison of number pairs, but in the “special case of ascending successive pairs” (p. 283), order might be recalled directly from memory. This explanation can account for faster response times for consecutive pairs as compared to non-consecutive pairs in order judgments in the present research. As shown in Table 3.2, the ordered stimuli in Triads 1 and 6 were consecutive sets of numbers, the ordered stimuli within Triads 2 and 5 were a mix of consecutive and non-consecutive numbers, and the ordered stimuli within Triads 3 and 4 were all non-consecutive numbers. Variability in participants’ use of successive magnitude comparisons versus serial search or direct access may account for differences across groups. These differences may also contribute to variability in whether SNARC effects were observed.

The reversal of the distance effect for order judgments as compared to magnitude judgments provides a possible explanation for the reversal of the SNARC effect for the English first-language participants in that participants may have used a mixture of procedures that were invoked based on the features of the stimulus set and the type of decision that was required. For example, if participants used magnitude comparisons on a substantial portion of the stimuli, then magnitude processes may have influenced performance more than order information. However, differential strategy use does not directly explain the reverse SNARC effect. The measure of SNARC is determined by the

differences in response times between right and left hands, not by the overall speed of response to consecutive versus non-consecutive numbers. The pattern of right minus left hand response times for the English-language participants could be an indication that they shifted their attention from left to the right on ascending sequences as they verified the order of the numbers in a serial search. This shifting of reference point may be similar to thinking about the face of an analog clock before making judgments about numbers (Bächtold et al., 1998), which can produce a reverse SNARC effect. In the descending block, participants may have processed the sequences from right-to-left (Bourassa, 2014) to determine order, given that all of the sequences that were ordered were descending in that block. This pattern of processing may support the typical SNARC effect. To determine if attention was shifted, physical measures such as eye movements might be useful, however, other factors such as stimulus composition may also be important and so Experiment 2 was designed to reduce the variability in the features of the sequences.

Modifications for Experiment 2

Overall, the results of Experiment 1 suggest that a SNARC effect may sometimes occur in order judgments, however, the data were somewhat noisy and only one group (Asian first language) showed typical SNARC effects across both ascending and descending ordered sequences. Turconi et al.'s (2006) findings, as well as the different patterns for consecutive versus non-consecutive sequences found in sequence judgment tasks suggest that the variability in sequence composition across triads may have contributed to instability in patterns of performance. Specifically, Lyons and Beilock (2009; see also Bourassa, 2014) found that participants responded to ordered sequential stimuli {1 2 3} more quickly than to ordered neutral stimuli {1 2 8} whereas the pattern

was reversed for unordered stimuli. Use of both sequential and neutral stimuli in Experiment 1 (Table 3.2) may have complicated participants' strategy selection and thus led to variability. This possibility can be seen in the differences among Triads for both percentage errors (Figure 3.1) and latencies (Figure 3.3). Triads 3 and 4 are made up of neutral stimuli whereas Triads 1 and 6 are sequential stimuli. The unequal distances between numbers within each stimulus set were also true to a lesser extent for Triads 2 and 5. As concluded by Turconi et al. (2006), decision making about stimuli of different inter-digit distances may result in different processes being used to determine order.

Accordingly, a modification to the stimuli was made in Experiment 2 in which equal inter-digit distances were used for all triads. This modification should result in more consistent and stable performance on the task. A second modification in Experiment 2 was to replace the keyboard with a response panel. The response panel, accurate to 1 ms, provides more accurate timing than the keyboard, which is only accurate to about 30 ms (Shimizu, 2002). SNARC effects are typically fairly small. Small inaccuracies in measurement could affect the reliability of the data. These two modifications should ensure more consistent performance across individuals and also reduce variability in response latencies. Third, participants' arithmetic fluency was assessed in Experiment 2 to test whether some of the variability in the present experiment was related to language group might be due to differences in numerical knowledge.

The power analysis at the end of Experiment 1 indicated that Experiment 2 should have at least 14 participants in each cell of the design to achieve a power of 80%. Because arithmetic skill was a factor, with a low and a high group, at least 28 participants were to be recruited for the experiment.

CHAPTER 4 – EXPERIMENT 2

In Experiment 1, SNARC effects were limited to the ordered stimuli. Further, a reverse SNARC effect was found for English first-language participants for the ascending task whereas a regular SNARC effect, with the usual negative slope, was found for the participants with an Asian first language for ascending and descending tasks. Both language groups read from left to right in their first language so these differences may not reflect language, *per se*. One explanation of the difference in direction of slope for the SNARC effects between the Asian- and English-language students is that performance may be related to arithmetic skill. Chinese-educated participants generally have higher levels of arithmetic skill than participants educated in Canada (Campbell & Xue, 2001; LeFevre & Liu, 1997; Sowinski et al., 2015). Thus, a plausible explanation for the differences between the language groups may be differences in arithmetic skill that influence patterns of performance on the sequence judgment task (Lyons & Beilock, 2011).

Some researchers have reported relations between arithmetic skill and the SNARC effect for traditional SNARC tasks. For example, Hoffman et al. (2014) reported a relationship between mathematical skill and the SNARC effect. The mean slope of the SNARC effect was steepest for the group with self-declared math difficulties and shallowest for those in math-heavy courses. This pattern is somewhat puzzling because it seems more likely that participants with stronger number skills would have more knowledge about numbers, including spatial relatedness and thus show larger SNARC effects. Of all the studies that have examined math experience or skill in relation to SNARC effects, however, all found either no significant relation (e.g., Dehaene et al.,

1993; Cipora & Nuerk, 2013) or, like Hoffman et al. (2014), a negative correlation (Fischer & Rottmann, 2005; Cipora et al., 2015).

Hoffman et al. (2014) provided a variety of reasons for the negative relation between arithmetic skill and the slope of SNARC effect. Those individuals who are better at arithmetic have more experience with numbers, which means more automatic access to number facts, which in turn means less working memory load during numerical tasks. Lyons and Beilock (2009) also reported faster performance on the order judgment task for individuals with better working memory. The lower load means that those better at arithmetic might be better able to inhibit irrelevant number information. In parity judgment tasks, the only relevant information is whether the number is odd or even and so the working memory load is moderate. Individuals with strong number representations may therefore show efficient recall of parity knowledge and inhibition of irrelevant spatial information, resulting the shallow slope for the SNARC effect for high-skill participants. In contrast, a slower, less efficient recall of parity of number combined with a lowered ability to inhibit the number's spatial position may result in a steeper slope for the lower-skill participants. In Experiment 2, participants completed a measure of arithmetic skill, the Calculation Fluency Test, to determine if arithmetic skill is related to the SNARC effect.

In Experiment 1, no SNARC effect was found for unordered Triads (i.e., stimuli made up of three single-digit numbers). However, the features of the stimuli were confounded with magnitude, such that the endpoint stimuli were counting series (e.g., 1 2 3 and 7 8 9) whereas stimuli in the centre were neutral (e.g., 1 2 6). In Experiment 2, only counting stimuli were used. Counting stimuli should result in maximal activation of

stored order information and thus maximally engage any spatial associations, such as the mental number line.

Method

Participants

Thirty-four students from Carleton University took part in the experiment in exchange for course credit. Data from three students who exceeded a 20% error rate were not analyzed. Five students read from right-to-left in their first language and so their data were not analyzed. Of the 26 remaining participants, 20 were female, and 23 were right-handed, with a mean age of 20.0 years ($SD = 2.9$). Cumulatively, the participants were fluent or conversant in 11 languages (Table 4.1). In contrast to Experiment 1, only two participants reported an Asian first language (Chinese).

Table 4.1. Experiment 2: Summary information of languages read by participants.

Reading Direction	<i>n</i>	First Language	Additional Languages
Left to Right	26	Bengali, Chinese (2), English, Ibo, Tamil, Yoruba	English, French, Hindi, Polish, Spanish, Twi, Yoruba
Right to Left	5	Arabic	English

Materials

Calculation Fluency Test. The Calculation Fluency Test was modeled after the French Kit of Cognitive Factors (French, Ekstrom, & Price, 1976). The Calculation Fluency Test (Sowinski et al., 2015) incorporates modified versions of the French Kit's addition, and of the subtraction/multiplication subtests. The test comprises one page each of: two-digit addition problems, two-digit subtraction problems, and two-digit by one-digit multiplication problems. These pages were preceded by five practice problems for

each operation type. Participants solved the problems, one row at a time, from left to right, without skipping any of the problems. Participants had one minute for each of the three problem pages. One point was assigned to each correct response. Total correct across all three pages was used as the index of skill.

Sequence Judgment Task. The stimuli, or Triads (Table 4.2) were constructed from the number set {1 2 3 4 5 6 7 8 9}. In the set of ordered Triads, only sequential numbers were used (i.e., counting sequences in Bourassa, 2014; sequential triads in Lyons & Beilock, 2009). Stimuli are shown in Table 4.2.

Table 4.2. Experiment 2: The single-digit number sets (Triads)

Task	Triad 1	Triad 2	Triad 3	Triad 4	Triad 5	Triad 6	Triad 7
Ascending							
ordered	1 2 3	2 3 4	3 4 5	4 5 6	5 6 7	6 7 8	7 8 9
unordered	1 3 2	3 2 4	3 5 4	5 4 6	5 7 6	7 6 8	8 7 9
Descending							
ordered	3 2 1	4 3 2	5 4 3	6 5 4	7 6 5	8 7 6	9 8 7
unordered	3 1 2	3 4 2	5 3 4	5 6 4	7 5 6	8 6 7	8 9 7

Equipment

The experiment was conducted using an IBM Pentium computer, an NEC MultiSync LCD 2060NX monitor, and the Cedrus RB-620 response panel. Stimuli and instructions were presented in the center of the monitor using black type on a white background. Participants sat at an approximate distance of 56 cm from the monitor. The stimuli, in 20 point Tahoma font, were made up of three single-digit numbers with three

spaces between each stimulus set taking up a width of 3.5 cm. The visual angle subtended 3.6°.

Procedure

Participants first completed the Calculation Fluency Test (CFT), followed by the Sequence Judgment Task. Participants were given one minute per page on the CFT. Six practice items are provided for each page.

In the sequence judgment task, participants judged whether the sequence of numbers (i.e., Triads) were in order. Seven ordered and seven unordered triads were presented within each block. The design included the within-groups variables of 2(Task: ascending, descending) x 2(Side of 'Yes' Response: left, right) x 7(Triad: 1 through 7) x 2(Stimulus Type: ordered, unordered) for a total of 56 distinct conditions. Each triad was repeated ten times within each condition, for a total of 560 trials.

Half of the participants began with the ascending task and the other half began with the descending task. Within each task, initial response key assignment was also counterbalanced, with half beginning with the left key assigned to yes, and the right key to no. In the second half of the experiment, response key assignment was reversed. Thus, the between-groups variables included 2(Task order: ascending/descending vs. descending/ascending) x 2(Side of 'Yes' response: left, right).

Participants were asked to judge whether the three single-digit numbers were in ascending (or descending) order. As in Experiment 1, participants performed six practice trials with feedback before the ascending task, and six practice trials with feedback before the descending task. If the response on a trial was incorrect, a notice appeared repeating the instructions, and the trial was repeated. The practice continued until all of the practice

trials had been answered correctly once. Participants were asked to be as fast and as accurate as possible. Between each task and key assignment participants were asked to take a break, resulting in three self-timed breaks. The experiment took approximately 30 minutes.

During the experiment, if a participant responded in less than 200 ms, a message appeared on the screen for two seconds asking the participant to “Please remember to be as accurate and as fast as possible”. On each trial, a mask made up of three asterisks (i.e., * * * appeared on the monitor for 1000 ms. They were replaced by the three single-digit number stimulus. The trial ended when the participant pressed one of the response keys.

Results

Data Trimming

Mean percentage of correct responses was 92.1%. Mean percent error responses ranged from 1.1% up to 20.7%, with a median percent error of 6.8%. After removing responses that were less than 200 ms or greater than or equal to three times the individual's mean response times, or duplicates of the preceding trial (as described in Experiment 1), 89.8% of the responses remained available for the analysis. Mean latencies were calculated for each condition in the 2(task: ascending, descending) x 2(side of response: left, right) x 2(stimulus type: ordered, unordered) design.

Speed-Accuracy Trade-off

A correlation was calculated (across all trials) between percentage errors and mean response time for all responses (correct and error). There were no speed-accuracy trade-offs, $r(24) = -.091$, $p = .659$. Speed-accuracy trade-offs were also investigated at the level of the two tasks, ascending and descending. There was no speed-accuracy trade-off

for the ascending task, $r(24) = .014$, $p = .945$, or for the descending task, $r(24) = -.149$, $p = .467$.

Calculation Fluency Test

The Calculation Fluency Test scores were used to categorize relative arithmetic skill within this experiment. An independent-samples t -test was conducted on mean fluency score. Experiment 2 mean fluency score was not significantly different from the mean fluency score for the Canadian subgroup from Sowinski et al. (2015), $t(266) = 0.68$, $p = .499$, indicating that the sample was representative of that population. The 50% of participants with the highest fluency scores ($n = 13$) were labeled as the high-skill group, and the 50% with the lowest fluency scores ($n = 13$) as the low-skill group. This resulted in two groups with different mean calculation fluency scores (Table 4.3).

Table 4.3. Experiment 2: Table of descriptive statistics for Calculation Fluency Test scores.

Group	<i>N</i>	Min	Max	<i>M</i>	<i>SD</i>	Median
All	26	6	65	29.7	13.52	31.5
Low	13	6	31	19.5	7.40	20.5
High	13	32	65	40.0	9.87	36.0

Error Analysis

The percentage errors were analyzed in a 2 (arithmetic skill: low, high) x 2 (Task: ascending, descending) by 2 (Stimulus type: ordered, unordered) by 7 (Triad 1 – 7) by 2 (Side of ‘Yes’ response: right, left) ANOVA with repeated-measures on the last three factors. A Greenhouse-Geisser epsilon adjustment of degrees of freedom was used if the assumption of homogeneity of variance was violated.

There was no main effect of Arithmetic skill (Table 4.4). There was a main effect of Task. Participants made fewer errors in the ascending task (6.9%, $SE = 0.9$) than in the descending task (9.0%, $SE = 1.1$). There was a main effect of Stimulus Type. Participants made fewer errors on ordered triads (6.5%, $SE = 0.8$) than on unordered triads (9.4%, $SE = 1.1$). There was a main effect of Triad and the linear trend was significant, $F(1, 24) = 27.28, p < .001, \eta_p^2 = .532$. Percentage errors generally increased from Triad 1 (4.5%, $SE = 0.79$) to Triad 6 (13.0%, $SE = 1.6$), however, percentage errors were higher on Triad 6 than on Triad 7 (9.1%, $SE = 1.5$). The interaction between Task and Triad was significant (Fig 4.1). Participants made more errors on the ascending than the descending task on Triads 2, 4, and 6. The number of errors made on ascending and descending tasks were the same for each of the other Triads.

Figure 4.1. Experiment 2: Interaction plot of Task by Triad for percentage of errors. The error bars are 95% CIs.

Table 4.4. Experiment 2: Source table of repeated measures ANOVA for percentage errors.

Source	<i>df</i> ₁	<i>df</i> ₂	<i>F</i>	<i>p</i>	η_p^2
Arithmetic Skill (low, high)	1	24	1.34	.259	.053
Task (ascending, descending)	1	24	11.93**	.002	.332
Task x Skill	1	24	1.91	.180	.074
Stimulus type (ordered, unordered)	1	24	20.25***	<.001	.458
Stimulus type x Skill	1	24	0.75	.397	.030
Triad (1 through 7)	3.7	89.2	10.23***	<.001	.299
Triad x Skill	3.7	89.2	0.54	.697	.022
Side	1	24	0.91	.348	.037
Side x Skill	1	24	7.14*	.013	.229
Task x Stimulus type	1	24	1.28	.270	.050
Task x Stimulus type x Skill	1	24	1.01	.326	.040
Task x Triad	4.7	112.9	3.09*	.014	.114
Task x Triad x Skill	4.7	112.9	2.13+	.070	.082
Stimulus type x Triad	4.4	104.4	2.07+	.084	.079
Stimulus type x Triad x Skill	4.4	104.4	1.43	.226	.056
Task x Stimulus type x Triad	4.2	99.7	3.58**	.008	.130
Task x Stimulus type x Triad x Skill	4.2	99.7	0.99	.420	.039
Task x Side	1	24	1.55	.225	.061
Task x Side x Skill	1	24	0.14	.714	.006
Stimulus type x Side	1	24	1.85	.187	.071
Stimulus type x Side x Skill	1	24	1.85	.187	.071
Task x Stimulus type x Side	1	24	1.71	.203	.067
Task x Stimulus type x Side x Skill	1	24	1.43	.244	.056
Triad x Side	4.7	112.7	1.37	.243	.054
Triad x Side x Skill	4.7	112.7	1.02	.407	.041
Task x Triad x Side	4.7	112.1	0.71	.608	.029
Task x Triad x Side x Skill	4.7	112.1	1.19	.321	.047
Stimulus type x Triad x Side	5.0	120.3	2.02+	.080	.078
Stimulus type x Triad x Side x Skill	5.0	120.3	0.87	.506	.035
Task x Stimulus type x Triad x Side	4.9	116.5	0.67	.642	.027
Task x Stimulus type x Triad x Side x Skill	4.9	116.5	1.05	.392	.042

+*p* = .08, * *p* < .05, ** *p* < .01, *** *p* < .001

The three-way interaction between Task, Stimulus Type, and Triad was significant (Fig 4.2). Triads 1, 6, and 7 for the ascending task showed better performance

for the ordered stimuli as opposed to the unordered stimuli whereas differences in performance for the descending sequences occurred only on Triads 5 and 7.

Figure 4.2. Experiment 2: Interaction plot of Task by Stimulus Type by Triad for percentage errors. The error bars are 95% CIs.

Figure 4.3. Experiment 2: Interaction plot of Stimulus Type by Triad for percentage errors (right – left hand). The error bars are 95% CIs.

The Stimulus Type x Triad x Side interaction was not significant (Fig. 4.3). Right-minus left-hand percentage errors was plotted. There were more errors made with right hand responses than left for Triad 2 on ordered sequences. There were marginally more errors made by right hand than left for Triad 6, also on ordered sequences. Otherwise both hands made the same number of errors across the triads for both Stimulus Types.

ANOVA on Mean Response Times

A 2(Gender: male, female) by 2(Starting button: left, right) univariate analysis was conducted on mean latencies. There were no significant effects. Hence, gender and starting button were collapsed as factors in the remaining analysis.

The mean correct latencies were analyzed in a 2(Task: ascending, descending) by 2(Stimulus type: ordered, unordered) by 7(Triad: 1 - 7) by 2(Side of response: left, right) by 2(Arithmetic skill: low, high) repeated-measures ANOVA, with Arithmetic skill a

between-subjects factor determined by a median split on the Calculation Fluency Test. The statistics for the analysis are shown in Table 4.5. The Greenhouse-Geisser epsilon adjustment of degrees of freedom was used if the assumption of homogeneity of variance was violated. The Sidak adjustment was used for multiple pair-wise comparisons.

There was a main effect of Task Direction. The participants judged sequences more quickly in the ascending task (904 ms, $SE = 41$) than in the descending task (1079 ms, $SE = 61$). There was a main effect of Stimulus type: participants responded faster to ordered triads (945 ms, $SE = 42$) than to unordered triads (1037 ms, $SE = 58$). There was a main effect of Triad and the linear trend of Triad was significant, $F(1, 25) = 63.95$, $p < .001$, $\eta_p^2 = .72$. Response latencies generally increased with Triad (Fig. 4.4). Triad 1 was responded to the most quickly of the seven triads.

Table 4.5. Experiment 2: Source table for the within-subject ANOVA on mean response time.

Source	<i>df1</i>	<i>df2</i>	<i>F</i>	<i>p</i>	η_p^2
Skill	1	24	2.80	.107	.104
Task (ascending, descending)	1	24	21.91***	<.001	.477
Task x Skill	1	24	0.06	.805	.003
Stimulus type (ordered, unordered)	1	24	10.69**	.003	.303
Stimulus type x Skill	1	24	0.36	.553	.015
Triad (1 through 7)	2.2	52.2	48.25***	<.001	.680
Triad x Skill	2.2	52.2	2.48+	.090	.093
Side	1	24	1.13	.299	.047
Side x Skill	1	24	2.06	.164	.079
Task x Stimulus type	1	24	3.28+	.082	.128
Task x Stimulus type x Skill	1	24	2.83	.105	.106
Task x Triad	4.1	97.9	2.56*	.042	.096
Task x Triad x Skill	4.1	97.9	1.75	.144	.068
Stimulus type x Triad	3.8	90.3	1.02	.399	.041
Stimulus type x Triad x Skill	3.8	90.3	1.68	.165	.065
Task x Stimulus type x Triad	3.0	71.7	2.17+	.099	.083
Task x Stimulus type x Triad x Skill	3.0	71.7	0.92	.437	.037
Task x Side	1	24	0.12	.736	<.01
Task x Side x Skill	1	24	1.19	.286	.047
Stimulus type x Side	1	24	<0.01	.998	<.01
Stimulus type x Side x Skill	1	24	0.70	.412	.028
Task x Stimulus type x Side	1	24	<0.01	.998	<.01
Task x Stimulus type x Side x Skill	1	24	1.33	.260	.052
Triad x Side	3.7	88.4	0.83	.499	.034
Triad x Side x Skill	3.7	88.4	0.90	.464	.036
Task x Triad x Side	4.0	96.0	1.04	.391	.042
Task x Triad x Side x Skill	4.0	96.0	0.61	.656	.025
Stimulus type x Triad x Side	4.1	97.5	0.70	.598	.028
Stimulus type x Triad x Side x Skill	4.1	97.5	0.89	.474	.036
Task x Stimulus type x Triad x Side	3.5	83.6	0.40	.784	.016
Task x Stimulus type x Triad x Side x Skill	3.5	83.6	1.07	.372	.043

+ $p < .10$ * $p < .05$, ** $p < .01$, *** $p < .001$.

Figure 4.4. Experiment 2: Line plot of mean response time for Triads. The error bars are 95%CIs.

The Task by Stimulus type interaction was significant. As shown in Figure 4.5, the difference in mean latencies between unordered and ordered stimuli was larger for the ascending Task (106 ms) than the descending Task (78 ms). The Task by Triad interaction was significant (Fig. 4.6). Latencies in the descending task were longer than those in the ascending task for each Triad, though the latency difference varied by Triad. The linear Task by Triad interaction was not significant. On ascending trials, there was no significant increase in latency from Triads 5 through 7, whereas for descending trials, latencies were faster on Triad 7 than on Triads 5 and 6.

Figure 4.5. Experiment 2: Interaction plot of mean response time for Task Direction by Stimulus type of stimulus set. The error bars are 95% CIs .

Figure 4.6. Experiment 2: Interaction plot of mean response time for the interaction of Task by Triad. The error bars are 95% CIs.

The interaction between Task Direction, Stimulus type, and Triad was significant (Fig. 4.7). Latencies on unordered trials were longer than on ordered trials across Triads for the ascending task. For Triads 6 & 7 the unordered sequences took significantly more time than the ordered sequences in the ascending task. This was not the case in the descending task. There were no latency differences between ordered and unordered sequences for Triads 6 & 7. For Triad 5 there was no significant difference in response time between ordered and unordered sequences for the ascending task, but there was for the descending task, with the ordered sequences having shorter latencies than the unordered sequences.

Figure 4.7. Experiment 2: Interaction plot of mean response times for Task by Stimulus type by Triads. The error bars are 95% CIs.

The interactions that included both Triad and Side of Response were not significant (Table 4.5). Thus, there was no evidence for SNARC effects in the overall ANOVA. In the next section, I explored SNARC effects using the approach that is common in many studies, that is, analyzing the difference in latencies to right versus left side responses.

A separate analysis was conducted on the first and the second half of the stimuli to determine if the presentation of ten blocks of the same stimuli would result in learning, and if so, if this learning would influence the SNARC effect. The mean correct latencies were analyzed in a 2(Task: ascending, descending) by 2(Stimulus type: ordered, unordered) by 7(Triad: 1 - 7) by 2(Side of response: left, right) by 2(Half: first half, second half of stimuli) by 2(Arithmetic skill: low, high) repeated-measures ANOVA, with Arithmetic skill a between-subjects factor determined by a median split on the Calculation Fluency Test. There were no significant Triad by Side of response interactions for either half of the stimulus presentations.

SNARC Effects

In Experiment 1, the SNARC effect varied with Language. I argued that the Language effect was most likely an effect of arithmetic skill, rather than of language per se. To compare SNARC effects for arithmetic skill, participants were categorised into high and low-skill groups using the median split of the Calculation Fluency Test results. Thirteen participants were in each group, with three males in each skill group.

Table 4.6. Experiment 2: Summary of the SNARC effect

Task+	Slope	% neg ^a	<i>t</i>	<i>df</i>	<i>p</i>
All sequences	-3.8	54	-1.43	25	.166
All ascending	-5.9	73	-2.28*	25	.031
All descending	-1.8	42	-0.45	25	.656
ascending ordered	-9.1	50	-2.53*	25	.018
ascending unordered	-2.9	43	-0.73	25	.470
descending ordered	0.6	36	0.13	25	.899
descending unordered	-3.2	43	-0.53	25	.599

^a Percentage of slopes that was negative across individuals.

* $p < .05$ (two-tailed)

Regressions were conducted using the difference in latencies (right- minus left-hand response time) as the dependent variable, and Triad as the independent variable. In Experiment 1, a significant reverse SNARC effect was found for the English-language group. As such, the single-sample *t*-test was conducted using a two-tailed test. In Experiment 2 the SNARC effect was significant for ascending sequence, but not for the descending sequence (Table 4.6). When the stimulus types were separated, the SNARC effect was significant for the ascending ordered sequences but not for ascending unordered sequences or either type of descending sequence. These findings support the hypothesis that the SNARC effect will be found for ordered, but not for unordered sequences.

The mean slopes and tests for significance are shown for each type of sequence for the two skill groups in Table 4.7. The only significant SNARC effect was found for the high skill participants on ascending ordered stimuli. Mean slopes were also negative for high skill participants on ascending, unordered stimuli; however, the mean slope was not significantly different from zero (although with a one-tailed test, it would be at $p =$

.046). Only 64% of the individual slopes were negative on unordered ascending trials whereas 79% of the individual slopes were negative for ordered stimuli. Despite a trend for negative slopes on all sequence types the pattern was less stable across individuals on unordered than on ordered stimuli.

In contrast, none of the slopes for the low-skill participants were significantly different from zero. Although the ascending, ordered slope was negative for the low-skill group, only 50% of the participants showed a negative slope, indicating that the effect was not consistent across individuals. Comparing the slopes for the Task and Stimulus Types for the two skill groups (Table 4.8) using independent-samples *t*-tests showed that there were no significant differences between low and high arithmetic skill ($ps > .1$).

Table 4.7. Experiment 2: Summary of the SNARC effect and the effects of Skill.

Skill	Task	Slope	% neg ^a	<i>t</i>	<i>df</i>	<i>p</i>
Low	ascending ordered	-5.3	50	-1.07	12	.308
	ascending unordered	2.4	43	0.39	12	.707
	descending ordered	3.7	36	0.46	12	.657
	descending unordered	-5.6	43	-0.67	12	.518
High	ascending ordered	-12.8	79	-2.51*	12	.028
	ascending unordered	-8.2	64	-1.84+	12	.091
	descending ordered	-2.4	43	-0.40	12	.695
	descending unordered	-0.9	43	-0.10	12	.922

^a Percentage of slopes that was negative across individuals.

+ $p = .091$, * $p < .05$ (two-tailed)

Table 4.8. Experiment 2: Summary of the comparison of SNARC effects between low and high Arithmetic skill groups for Task and Stimulus Type.

Task	Arithmetic Skill					
	Low		High		Low vs. High	
	Slope	<i>SD</i>	Slope	<i>SD</i>	<i>t</i>	<i>p</i>
ascending ordered	-5.3	18.1	-12.8*	18.4	1.04	.31
ascending unordered	2.4	22.7	-8.2	16.0	1.38	.18
descending ordered	3.7	29.6	-2.4	21.9	0.61	.55
descending unordered	-5.6	30.2	-0.9	32.8	-0.38	.71

* significant at $p < .05$ (two tailed)

Figure 4.8 and Figure 4.9 illustrate the patterns of SNARC effects across Task, Arithmetic skill groups and Stimulus Types. The significant SNARC effect for the high Arithmetic skill group for the ascending ordered triads followed the traditional shape with a negative slope and the data points are orderly (Fig. 4.8B).

Figure 4.8. Experiment 2: Line plot of the difference in mean response times by Ascending by Stimulus type by Triad by Arithmetic skill group

Figure 4.9. Experiment 2: Line plot of the difference in mean response times by Descending by Stimulus type by Triad by Arithmetic skill group

The data for the unordered trials is less stable, even though it is generally demonstrates a flat to negative slope (Fig. 4.8C & Fig. 4.9G, & H). In contrast, there were no SNARC effects for the low Arithmetic skill group. The difference in mean response times for the ordered (Fig. 4.8A) indicated that the right hand was predominantly slower than the left for the ascending task, but faster than the left for the descending task (Fig 4.9E). The SNARC effect for the ascending unordered triads for the

high arithmetic skill group was trending towards significance (Table 4.7). The percentage of slopes that were negative across the participants was less than 70%, indicating individual variability in judging the stimuli. The data from Figure 4.8D was less stable than Figure 4.8B, but was more stable than the data in Figure 4.8 and Figure 4.9.

Correlations between calculation fluency and sequence slopes.

As shown in Table 4.9, none of the correlations between calculation scores and tasks were significant. The correlation between the slopes of the descending unordered and the ascending ordered judgments approached significance.

Table 4.9. Experiment 2: Correlations among SNARC coefficients and Calculation Fluency scores.

	CFT	Asc-Ord	Asc-Un	Des-Ord
1. Calculation fluency scores (CFT)	-			
2. Ascending ordered	-.06	-		
3. Ascending unordered	-.14	-.04	-	
4. Descending ordered	-.01	.02	-.06	-
5. Descending unordered	.12	.37+	.17	-.10

Note. $n = 26$; $+ p = .063$.

The slopes increased together. Looking at Figure 4.10 there is an outlier with a very low descending unordered value. Removing the data from the one participant results in the correlation no longer being significant, or near significance ($r = .33$, $p = .111$). The correlation is not stable.

Figure 4.10. Experiment 2: Scatterplot of slopes for descending unordered and ascending ordered. Line is line of best fit ($n = 26$).

Power Analysis

A post hoc power analysis, using MorePower 6.0.1 (Campbell & Thompson, 2012) was conducted on the SNARC effects (Appendix A, Table A.2). Power for the significant SNARC effects (Table A.3) ranged from .37 up to .63. Using the standard deviation (18.4) and η^2 (.34) from the significant effect (but not the marginal effect) there should be at least 18 people in each cell of the next experiment.

Discussion

The purpose of Experiment 2 was to determine if there was a space-number association for both ordered and unordered stimuli. I used different stimuli than in Experiment 1, focussing on those stimuli with a distance of 1 between the digits of each triad (counting stimuli). A second question was whether the findings from Experiment 1

arose out of arithmetic skill rather than language. In Experiment 2, the SNARC effect was found only in the ascending task for the high Arithmetic skill group on ordered stimuli. Had there been a significant SNARC effect for the unordered stimuli, this would have provided strong evidence that the effect was driven by Magnitude, as the small number set {3 1 2} employs the same space on the left side of the number line as the small number set {1 2 3}. The finding that only ordered sequences resulted in a SNARC effect for the order judgment task suggests that magnitude is not the sole arbiter of the effect. Minimally, therefore, order and magnitude, not just magnitude, appear to influence relations between number and space as reflected in SNARC effects.

The finding of a SNARC effect for the high-skill but not for the low-skill group does not parallel previous findings for SNARC effects in parity tasks (Cipora et al., 2015; Dehaene et al., 1993; Fischer & Rottmann, 2005; Hoffman et al., 2014). In those experiments, the tendency was for individuals with lower levels of skill or less experience with mathematics to show SNARC effects (i.e., a larger negative slope). This difference between the parity and sequence judgment tasks leads to the question whether the present finding is actually a typical SNARC effect. This question was directly tested in the next experiment with the addition of a parity task. If the space-number associations for the Parity and the Sequence Judgment Tasks are related, this will be seen as evidence that the space-number effect for the Sequence Judgment Task was a SNARC effect.

Limitations of Experiment 2

One limitation of the present research was that the sample size of 26 was modest. Although Wood et al. (2006) determined that 24 participants were needed for a power of 83%, there were only 13 individuals in each skill category in the present experiment,

hence reducing the power to detect group differences. Therefore, a larger sample size was used in Experiment 3.

Experiment 3 was designed to replicate the number-space effect found in Experiments 1 and 2, and test some of the possible reasons why the effect may be more elusive than in more traditional tasks where the SNARC has been observed. In Experiments 1 and 2, judging the unordered sequences may not have activated a spatial representation relative to the full number set 1 through 9. The relevant mental reference for an unordered sequence may be each number's positioning within the three numbers of the stimulus (Triad) as opposed to the set's positioning within the full set of numbers. In Experiment 3, some participants learned the relative ordering of the unordered Triads both relative to the full set of numbers {1-9} and to the other number sets (Triads).

Participants in the training condition performed two tasks with the unordered sequences that were typed on index cards. They first located each Triad by pointing to the Triad's position on a number line. They then sorted the set of unordered sequences from left to right from lowest to highest number set. For example, when sorted, the index card with {1 3 2} would be to the left of the index card with {2 4 3}, and the index card containing {7 9 8} would be farthest to the right. It was hypothesised that the training on relational ordering of Triads will enable the SNARC effect for the unordered Triads.

The power analysis at the end of Experiment 2 indicated that Experiment 3 should have at least 18 participants in each cell of the design to achieve a power of 80%. As arithmetic skill was again a factor, with a low and a high group, at least 36 participants were needed.

CHAPTER 5 – EXPERIMENT 3

In both Experiments 1 and 2, some participants showed SNARC effects on ordered sequences, specifically, Asian participants in Experiment 1, and those with higher levels of arithmetic skill in Experiment 2. However, there were no significant SNARC effects in either experiment for unordered stimuli. Accordingly, there was no clear evidence that magnitude is the impetus for SNARC but the results were also not strong for ordered sequences. In Experiment 1, a spatial numerical association was found for both ascending and descending tasks. In Experiment 2, using stimuli only made up of sequential digits, a spatial numerical association was only found for the ascending task. As the spatial numerical association was found in both experiments for the ascending task, participants were only asked to judge ascending sequences in Experiment 3.

In the present experiment, participants learned the global ordering of the unordered Triads to determine if a SNARC effect could be found for unordered sequences. The sample size was also increased dramatically in an attempt to get more stable results. Furthermore, the SNARC effect was tested with the traditional task, parity judgments, with the same participants, to determine whether the apparatus and participant group used in the current research would show typical patterns of performance in the parity task. The parity task is the most common context wherein SNARC effects have been observed: In the meta-analysis by Wood et al. (2008), 50% of the experiments analyzed used parity judgments to test for SNARC effects.

A number of experiments that tested for SNARC effects have been conducted wherein an order was learned or the context changed. For example, when participants first had to think about the face of a clock and its numbers a reverse SNARC effect was

observed, presumably because when the numbers on the face of a clock are unfolded the large numbers are on the left and the small numbers are on the right (Bächtold et al., 1998). Prado et al. (2008) informed participants about where some of the guests were sitting in relation to other guests at a dinner party. The participants worked out and identified the full seating order at the table. Prado et al. found that when participants made judgments about the seating their data showed a SNARC effect. In Experiments 1 and 2, participants may have processed ordered sequences relative to the full number set $\{1 - 9\}$ and so such sequences as $\{1\ 2\ 3\}$ may have activated more of the number line, for example, through continued activation of the sequence $\{e.g.,\ 4\ 5\ 6\ \dots\}$. In contrast, the unordered sets do not predict the next sequence and so may not have resulted in as much activation of the whole number line from 1 to 9. To influence order activation, participants sorted card sets of the unordered stimuli and located the position of the Triad on a physical image of a number line to emphasise the order of the unordered Triads relative to the other unordered Triads. By increasing the relevance of the mental positioning of each number set relative to the others, a SNARC effect may be found for unordered sequences. A control group performed a Number Line Placement Task (they located single digits on a number line) in place of the sorting of the unordered Triads.

The selection of unordered stimuli was controlled more closely in this experiment as compared to the previous two. Notably, unordered stimuli follow differing patterns of relative relations among the numbers. For example, starting with the ordered sequence 2 3 4, there are four different unordered sequences: 2 4 3 and 3 4 2, where there is an increase between the first two digits, and 4 2 3 and 3 2 4 where there is a decrease between the first two digits. Bourassa (2014) showed that participants use a strategy of

sequential comparisons in the ascending task, such that they more quickly reject the sequences where there is an initial decrease than the ones where there is an initial increase. Allowing participants to adopt different strategies may add variability to the responses. Accordingly, in the present experiment, only unordered sequences of the increasing type were used.

Even within this constrained set of stimuli, it is possible that sequences such as 2 4 3, where the initial difference is greater than one, may be rejected more quickly than sequences such as 3 4 2, which are more similar to a possible ordered sequence because the first two digits are different by one (i.e., 3 4 5). Thus, the two types of unordered sequences were compared in the current experiment. Sequence type included three levels, ordered, unordered-1 (e.g., 2 3 1), and unordered-2 (e.g., 1 3 2). If participants develop a strategy related to the differences between first two digits, they may be able to reject unordered-2 sequences faster than unordered-1 sequences, potentially reducing the opportunity for a SNARC effect to be observed.

Method

Participants

Participants were recruited from two groups: Carleton University's online experimental system, and from a list of off-campus volunteers known to the researcher. All participants spoke English. In their first language, 73 individuals read and wrote from left to right, and five read and wrote from right to left (Table 5.1). Cumulatively, the participants were fluent or conversant in 17 languages. Because of the small number of right to left readers ($n = 5$), and the expectation that the typical SNARC effect follows the direction of reading, their data was not used in this analysis. The first fifty-nine

participants were assigned to the experimental group. The nineteen participants that followed were assigned to the control group.

Of the remaining 73 participants, 66 were right-handed and 40 were female with a mean age of 40.5 years ($SD = 20.2$), ranging from 17 to 79 years of age. Thirty-seven took part on-campus, and 36 took part off-campus. Carleton University students were offered the option of course credit or \$10. Off-campus participants received \$10.

Table 5.1. Experiment 3: Summary information of languages read by participants (in alphabetical order).

Reading Direction	n	First Language	Additional Languages
Left to Right	73	Chinese, English, French, Ibo, Malay, Polish, Spanish	Chinese, Dutch, English, French, German, Greek, Ibo, Latin, Mohawk, Spanish
Right to Left	5	Arabic, Farsi, Persian	English, French, Russian, Turkish

Equipment

A Pentium computer, a MultiLink LCD monitor, and a Cedrus RB-620 response panel were used for both the Parity Judgment Task, and the Sequence Judgment Task. The stimuli were controlled and the results collected using Superlab version 4.5 software. All instructions and stimuli were presented in Tahoma, regular, 20-point font, centered on the screen. When responding on the computer, participants sat at a self-defined comfortable distance from the monitor, approximately 50 cm. The width upon the monitor of the three single-digit numbers was 3.5 cm, subtending 3.6° .

The Number Line Placement Task performed by the control group was done using the EstimationLine application on an iPad (Hume, 2011).

Materials

Participants from the experimental group completed four activities: a Calculation Fluency Test (Sowinski et al., 2014), a Parity Judgment Task, a Training Task, and a Sequence Judgment Task. In place of the training task participants from the control group completed a Number Line Placement Task.

Calculation Fluency Test. The same test was used as in Experiment 2.

Parity Judgment Task. Single-digit stimuli for the Parity Task were drawn from the number set $\{1,2,3,4,5,6,7,8,9\}$.

Number Line Placement Task. Single-digit stimuli were drawn from the number set $\{1,2,3,4,5,6,7,8,9\}$.

Sequence Judgment Task. The stimuli were drawn from the number set $\{1,2,3,4,5,6,7,8,9\}$. Only ascending sequences were used. Each ordered stimulus (Triad) was made up of three numbers with a distance of 1 (Table 5.2). The two unordered sequences were inversions of the ordered set as per the algorithm (a b c $\Rightarrow$ b c a | a c b) such that the first two numbers in the sequence were in an ascending order (Table 5.2). The unordered-1 stimulus sets had a distance of one between the first two digits. The unordered-2 stimulus sets had a distance of two between the first two digits. The stimuli were presented in black on a white screen in Tahoma font size 20.

Table 5.2. Experiment 3: Single-digit number sets (Triads).

Type	1 st	2 nd	3 rd	4 th	5 th	6 th
ordered	1 2 3	2 3 4	3 4 5	5 6 7	6 7 8	7 8 9
unordered-1	2 3 1	3 4 2	4 5 3	6 7 5	7 8 6	8 9 7
unordered-2	1 3 2	2 4 3	3 5 4	5 7 6	6 8 7	7 9 8

Training Task. Three sets of index cards were created for each of the two unordered sequences. The image of a number line with the numbers 1 and 9 marked at the two end points were printed in black on an 8 ½ by 11 sheet of white paper. Only ascending unordered sequences were used. Triads (4 6 5) and (5 6 4) were excluded from the Training task to reduce the number of stimuli presentations to the participants.

Procedure

All participants wrote the Calculation Fluency Test, and performed both the Parity Task and the Sequence Judgment Task. Fifty-nine participants performed the Training Task before the Sequence Judgment Task (experimental group). Nineteen participants performed a Number Line Placement Task instead of the Training Task (control group).

Participants began with the Calculation Fluency Test. After the Calculation Fluency Test all participants completed the Parity Task. On each trial, participants decided if a single-digit number was odd or even. Participants were instructed to be both fast and accurate. An asterisk * appeared on screen for 1000 ms. The asterisk was replaced by a single-digit number which remained on the screen until the participant pressed a button on the Cedrus response panel. As in Experiments 1 and 2, participants performed six practice trials with feedback before the ascending task, and six practice trials with feedback before the descending task. If the response on a trial was incorrect, a

notice appeared repeating the instructions, and the trial was repeated. The practice continued until all of the practice trials had been answered correctly once. Half of the participants began by pressing the left button to indicate Odd during the practice component and the first half of this task; half began with the left button indicating Even. Halfway through the experiment, this button assignment was reversed for both the practice component and the task. There were seven blocks of nine stimuli repeated for two sides of response, resulting in 126 trials.

A Training Task followed the Parity Task. The purpose of the new task was to induce an order between the sets of unordered sequences for the experimental group. Participants were shown six cards in random order. The participants were asked to show where the numbers on the index card, as a set, could be found by pointing to one location on the printed number line. This show and point was done twice for the set of cards. The participants then sorted the six index cards into order from smallest set to largest, from left to right. Half the experimental group were shown the unordered-1 set of numbers (a distance of 1 between the first two numbers of the three number sequence). The other half of the experimental group saw the unordered-2 sets of numbers (a distance of two between the first two numbers of the three number sequence). No cards were shown for the ordered stimuli. No feedback was required because all participants correctly pointed to the location of all of the triads on the number line. Also, they all correctly sorted the triads in relative order.

The control group were shown nine single-digit numbers {1,2,3,4,5,6,7,8,9} in random order on an iPad. They responded by tapping the location of each number on a number line presented on the iPad screen. No measurements were recorded.

The sequence judgment task was the same as that used in the previous experiments. Half of the participants began with the ‘yes’ response being assigned to the left button and ‘no’ to the right button. Halfway through the experiment assignment was reversed. Assignment of participant to condition was random.

As seen in Table 5.2, there were six Triads for each of the three Sequence Task Types (ordered, unordered-1, unordered-2). Seven blocks of these 18 stimuli were presented. This was repeated twice for the two response button assignments (left = yes, right = yes) resulting in 84 ordered and 168 unordered presentations of stimuli, a total of 252 responses. Participants were told to be both fast and accurate. The experiment took approximately 35 minutes.

Results

Skill Groups

The Calculation Fluency Test scores were used to categorize relative arithmetic skill within this experiment. An independent-samples *t*-test was conducted on mean fluency score. Experiment 2 mean fluency score was significantly greater than the mean fluency score for the Canadian subgroup from Sowinski et al. (2015), $t(254) = -4.11, p < .001$. The 50% of participants with the highest fluency scores ($n = 36$) were labeled as the high-skill group (Table 5.3). They scored significantly lower than the Chinese subgroup from Sowinski et al. (2015), $t(92) = -2.20, p = .030$. The 50% with the lowest fluency scores ($n = 37$) were labeled as the low-skill group. The low-skill arithmetic score was not significantly different from the score for the Canadian subgroup from Sowinski et al. (2015), $t(278) = 1.55, p = .122$.

Table 5.3. Experiment 3: Table of descriptive statistics for the Calculation Fluency Test.

Group	<i>N</i>	Minimum	Maximum	<i>M</i>	<i>SD</i>	Median
All	73	11	86	39.2	14.8	38.0
Low	37	11	37	28.1	6.77	29.0
High	36	38	86	50.0	12.20	46.5

Sequence Task

Data Trimming.

Mean percentage error was 3.8%, ranging from a low of 0.4 to a high of 20.6 with a median percentage error of 3.2%. After removing responses that were less than 200 ms or greater than or equal to three times the individual mean response times, and duplicates, 94.0% of the responses were retained for analyses. Data from all 73 participants were used in the Sequence Task analysis.

Speed-Accuracy Trade-off.

There was no correlation between percentage errors and mean correct response time, $r(71) = .039$, $p = .746$. As such, there was no speed-accuracy trade-off.

Error Analysis.

The percentage errors were analyzed in a 3(Stimulus type: ordered, unordered-1, unordered-2) by 6(Triad 1 – 6) by 2(Side of ‘Yes’ response: right, left) by 2(Arithmetic Skill: low, high) mixed ANOVA with Arithmetic Skill as a between-subjects factor. A Greenhouse-Geisser epsilon adjustment of degrees of freedom was used if the assumption of homogeneity of variance was violated. The statistical analysis is shown in Table 5.3.

There was a main effect of Arithmetic skill. Participants in the low-skill group made more errors (4.8%, $SE = .56$) than those in the high-skill group (2.7%, $SE = .57$). There was a main effect of Triad. Participants made significantly fewer errors on Triads 1, 2, and 3 (2.1, 2.3, and 2.7%) than on Triads 5 (7.1 %) and 6 (4.9 %). Percentage errors for Triad 4 (3.4%) were significantly lower than those on Triad 5, but not Triad 6. The linear trend of Triad on percentage errors was significant, $F(1,71) = 46.13, p < .001, \eta^2 = .394$, indicating that, in general, errors increased with number magnitude.

There was a main effect of Stimulus Type on percentage errors. Participants made fewer errors on unordered-1 sequences (2.4%, $SE = .040$) than on ordered (4.7%, $SE = 0.47$) or unordered-2 sequences (4.1%, $SE = 0.52$), which were not different. This patterns was unexpected, in that unordered-1 sequences (e.g., 2 3 1) should have been harder to reject than the unordered-2 sequences (e.g., 1 3 2) if participants had used a strategy that considered the difference between the first two digits.

Table 5.4. Experiment 3: Source table of repeated measures ANOVA for percentage errors.

Source	<i>df1</i>	<i>df2</i>	<i>F</i>	<i>p</i>	η_p^2
Arithmetic skill (low, high)	1	71	6.76*	.011	.087
Stimulus Type (ordered, unordered-1, -2)	1.9	136.5	15.69***	<.001	.181
Stimulus Type x skill	1.9	136.5	3.23*	.044	.044
Triad (1 through 6)	2.7	190.3	15.04***	<.001	.175
Triad x skill	2.7	190.3	0.48	.674	.007
Side (left, right)	1	71	0.14	.710	.002
Side x skill	1	71	0.68	.413	.009
Stimulus Type x Triad	4.9	348.5	6.01***	<.001	.078
Stimulus Type x Triad x skill	4.9	348.5	0.85	.514	.012
Stimulus Type x Side	1.4	96.7	0.95	.359	.013
Stimulus Type x Side x skill	1.4	96.7	1.95	.161	.027
Triad x Side	4.6	326.4	0.07	.996	.001
Triad x Side x skill	4.6	326.4	0.24	.937	.003
Stimulus Type x Triad x Side	7.9	560.5	1.21	.291	.017
Stimulus Type x Triad x Side x skill	7.9	560.5	0.65	.733	.009

* $p < .05$, *** $p < .001$

The main effect was qualified by the interaction of Stimulus Type and Triad (Fig 5.1). It appears that one particular ordered-2 sequence was more difficult to reject (e.g., 6 8 7) than the unordered-1 pair (e.g., 7 8 6), because there were no differences between any other pair of unordered sequences. The interaction between Stimulus Type and Arithmetic skill was significant (Fig. 5.2). Low-skill participants made more errors than high-skill on ordered sequences and on unordered-2 sequences whereas they did not differ on unordered-1 sequences.

Figure 5.1. Experiment 3: Interaction plot of Stimulus Type by Triad for mean percentage errors. The error bars are 95% CIs.

Figure 5.2. Experiment 3: Interaction plot of Stimulus Type by Arithmetic skill for mean percentage errors. The error bars are 95% CIs.

In summary, despite an overall low percentage of errors, the task was more difficult for the low than the high arithmetic skill group. The unordered-1 sequences were more difficult than the unordered-2 stimuli, but only for the high skilled group, and only because the high skill group had made so few errors on unordered-1 stimuli. The low skill group made more errors on the ordered stimuli than the two unordered sequence types. This is initially surprising as judging ordered sequences would presumably be the easier than judging unordered sequences. One reason for this may be the experimental design. Each participant saw 84 ordered and 168 unordered triads, an unbalanced presentation of stimuli. With more unordered than ordered triads the low skill group may have been biased towards responding unordered. This biasing would have lead to the higher number of errors.

Response Time Analysis.

A 2(Gender: Male, Female) by 2(Starting button: Left, Right) by 2(Location: on campus, off campus) univariate ANOVA was conducted on mean latencies. There were no significant effects. Hence, gender, starting button, and location of experiment were collapsed as factors in the remaining analysis.

Mean latencies were analyzed in a 3(Stimulus type: ordered, unordered-1, unordered-2) x 6(Triad: 1 - 6) x 2(Side of 'Yes' response: right, left) by 2(Arithmetic Skill: low, high) mixed ANOVA with Arithmetic skill as a between-subjects factor (Table 5.5). As in previous experiments, Greenhouse-Geisser epsilon adjustment of degrees of freedom and Sidak adjustments were used when homogeneity of variance was violated.

There was a main effect of Stimulus type. Participants were faster to judge ordered sequences (919 ms, $SE = 26$) than either unordered-2 sequences (958 ms, $SE = 27$) or unordered-1 sequences (968 ms, $SE = 28$), which were not significantly different.

Table 5.5. Experiment 3: Source table for the analysis of mean response times in the Sequence Judgment Task.

Source	df_1	df_2	F	p	η^2
Arithmetic skill (low, high)	1	71	10.78**	.002	.132
Stimulus type (ordered, unordered-1, unordered-2)	1.8	130.0	14.27***	<.001	.167
Triad (1 – 6)	3.5	246.1	111.10***	<.001	.610
Side (left, right)	1	71	0.75	.389	.010
Skill by Stimulus type	1.8	130.0	1.62	.204	.022
Stimulus type x Triad	7.9	557.4	12.61***	<.001	.151
Stimulus type x Side	1.4	97.2	0.28	.669	.004
Skill x Triad	3.5	246.1	3.77**	.008	.050
Triad x Side	4.2	297.1	0.37	.837	.005
Side x Skill	1	71	0.07	.792	.001
Stimulus type x Triad x Skill	7.9	557.4	1.64	.112	.023
Stimulus type x Side x Skill	1.4	97.2	0.48	.550	.007
Triad x Side x Skill	4.2	297.1	0.57	.693	.008
Stimulus type x Triad x Side	7.4	524.4	1.34	.225	.019
Stimulus type x Triad x Side x Skill	7.4	524.4	1.42	.190	.020

** significant at $p < .01$, *** significant at $p < .001$

Figure 5.3. Experiment 3: Line plot of Triad by mean Response Time. The error bars are 95% CIs.

Figure 5.4. Experiment 3: Interaction plot of Arithmetic skill by Triad by mean Response Time. The error bars are 95% CIs.

There was a main effect of Triad (Fig. 5.3). The linear trend for Triad was significant, $F(1,71) = 290.09, p < .001, \eta_p^2 = .803$. The response latencies generally increased with Triad. There was an end point advantage, with participants responding to Triad 6 faster than to Triad 5.

There was a Triad by Arithmetic skill interaction (Fig. 5.4). The linear trend of the interaction was not significant, $F(1, 69) = 1.18, p = .282, \eta_p^2 = .017$; both groups showed increases as a function of Triad. Instead, the interaction arises because the low-skill group were relatively faster on Triad 1 as compared to Triad 2, whereas the high-skill group did not show a significant difference between these two Triads. In other respects, the pattern of response times was similar across Triads for the two groups.

Figure 5.5. Experiment 3: Interaction plot of Stimulus type by Triad for mean Response Time. The error bars are 95% CIs.

There was a Triad by Stimulus type interaction (Fig 5.5). The interaction was mainly due to a large advantage for ordered Triad 1 (i.e., 1 2 3). There was also a small disadvantage for unordered-2 Triad 5 (i.e., 6 8 7) and for unordered-1 Triad 6 (i.e., 8 9 7). However, there were no Stimulus type differences in response time for Triads 2, 3, and 4.

In summary, there were effects of stimulus type, skill, and Triad in this experiment, but no significant effects involving side of response. Despite the larger sample size and the more controlled stimuli, the SNARC effect continued to be elusive.

SNARC effect for Sequence Judgments.

The next step of the analysis was to investigate the SNARC effect directly, using the between-subjects factor of Skill. The SNARC effect arises out of the interaction between stimulus (Triad) and Side of response. Although the interaction between Triad and Side of response was not significant in the main analysis (Table 5.5), recall that participants in the experimental group ($n = 57$) took part in a card sorting activity. Before performing the Sequence Task, 31 sorted the unordered-1 Triad cards and 26 sorted the unordered-2 Triad cards.

Ordered trials. Single-sample *t*-tests were conducted on slope values for the ordered stimuli for the whole sample and for the low- and high-skill groups separately, as shown in Table 5.6. As no training had been done on the ordered stimuli, the slopes for all of the participants were used in the analysis. There was a SNARC effect for ordered Triads for participants in the high-skill group (Fig. 5.6), but not for those in the low-skill group (Fig. 5.6). This pattern replicates the results of Experiment 2.

Table 5.6. Summary of the ordered Stimulus type SNARC effects by Skill group.

Group	Stimulus Type	Slope	% neg ^a	<i>t</i>	<i>df</i>	<i>p</i>
All participants	ordered	-6.1	58.9	-1.58	72	.119
Low Skill	ordered	0.4	51.3	0.62	36	.951
High Skill	ordered	-12.8	66.7	-2.71**	35	.010

^a Percentage of slopes that was negative across individuals.

** $p < .01$ (two-tailed)

Figure 5.6. Experiment 3: Scatterplot of ordered Triads by difference in mean response time by Arithmetic skill.

Unordered trials. The Training Task was analysed for each unordered type and Skill group. Recall that during the Training Task, 26 participants sorted index cards with the unordered-1 Triads printed on them and 31 sorted index cards with the unordered-2 Triads printed on them. There were no significant spatial-numeric associations for the unordered stimuli (Table 5.7), although the effect for unordered-1 sequences approached significance for the low-skill group trained with those sequences.

Table 5.7. Experiment 3: Summary of the unordered Stimulus type SNARC effects by Training Group.

Skill	Training Cards ^a	Stimulus	<i>M</i> (Beta)	% neg ^b	<i>t</i>	<i>df</i>	<i>p</i>
Low	Type 1	unordered-1	-12.2	79	-1.76+	18	.095
		unordered-2	6.2	32	0.79	18	.439
	Type 2	unordered-1	5.8	30	0.68	9	.511
		unordered-2	7.3	40	0.98	9	.354
High	Type 1	unordered-1	4.6	42	0.74	11	.474
		unordered-2	10.2	42	1.04	11	.319
	Type 2	unordered-1	5.3	69	1.00	15	.332
		unordered-2	2.9	31	0.45	15	.659

+ $p = .095$ (two-tailed)

^a Type 1 cards matched unordered-1 Triads. Type 2 cards matched unordered-2 Triads.

^b Percentage of slopes that were negative across individuals.

Table 5.8. Experiment 3: Summary of the unordered Stimulus type SNARC effects for the control group.

Skill	Stimulus	<i>M</i> (Beta)	% neg ^a	<i>t</i>	<i>df</i>	<i>p</i>
All	unordered-1	-2.2	50	-0.20	15	.846
All	unordered-2	-1.6	56	-0.32	15	.750
Low	unordered-1	-4.9	38	-0.23	7	.830
	unordered-2	0.6	38	0.07	7	.947
High	unordered-1	0.4	62	0.05	7	.962
	unordered-2	-3.8	75	-0.67	7	.529

^a Percentage of slopes that was negative across individuals.

Controls, and Unordered trials. An analysis of the participants who did not take part in the training task of sorting cards was conducted next (Table 5.8). There were no

significant SNARC effects. There was a large slope for the unordered-1 stimuli for the low skill group. An investigation of the individual slopes for those stimuli revealed one participant had a very large slope that influenced the overall mean, as confirmed by the low percentage of slopes that were negative for the low skilled group.

Power analysis for Sequence task

A post hoc power analysis, using MorePower 6.0.1 (Campbell & Thompson, 2012) was conducted on the SNARC effects (Appendix A, Tables A.3 – A.5). Power for the significant SNARC effect for the ordered sequences was .75. Power for the marginally significant SNARC effect for the experimental group who sorted cards (low skill, unordered-1) during the training task was .37. Power for the unordered sequences for controls are listed in Table 5.15.

Discussion of SNARC for sequence judgments.

SNARC effects were found for the ordered stimuli, but only for the high Arithmetic skill participants (Table 5.6). It appears that the individuals who had higher scores on the Calculation Fluency Test might maintain a better spatial representation of numbers, or of patterns of numbers, than the participants whose scores were on the lower half of the test.

Participants underwent a Training task wherein they learned the order of the Triads. The training was intended to provide participants with a global referencing of the Triad relative to the other Triads, rather than the local referencing of the numbers to each other within the Triads. The spatial numerical associations were not significant for unordered trials.

Did the lack of a SNARC effect amongst participants who scored on the bottom half of the Calculation Fluency Test occur because the low arithmetic skill participants did not activate a mental representation of numbers? To test for this participants also undertook the Parity Judgment Task.

Parity Task

Errors.

Two participants did not perform the experiment as directed. Their data were excluded. Seventy-one participants each provided 126 responses resulting in 8946 responses. The percentage correct response was 96.0% with median percentage errors 3.2%, ranging from 0% to 16.7%. The stimuli were presented in random order from the list of 9 numbers using Superlab 4.5. After each iteration of the list it was possible the same number could appear twice in a row (end of the first list, beginning of the next list). The response time after a repetition may then be speeded up if the correct answer is retained in working memory, or it might be slowed down if the participant believes they already responded to the stimulus. A repetition introduces a possible difficulty in determining if an extra factor has been included in the experiment. After removal of errors, extreme values (either less than 200 ms, or greater than the mean response time plus three standard deviations), and the second response of a duplicate presentation of the same stimulus (which ranged from one to four trials per participant), 94.4% of the responses were used in the analysis.

Speed-Accuracy Trade-off.

The correlation between number of errors and mean Response Time for each participant was not significant, $r(70) = .036, p = .781$. As such, there was no speed-accuracy trade-off.

Error Differences.

The percentage errors were analyzed in a 9(Number: 1-9) by 2(Side of response: right, left) by 2(Arithmetic skill: low, high) mixed ANOVA. Arithmetic skill was a between-subjects factor. A Greenhouse-Geisser epsilon adjustment of degrees of freedom was used if the assumption of homogeneity of variance was violated. As shown in Table 5.9, there was a significant main effect of Number on percentage error. The linear effect was not significant however, $F(1, 70) = 1.46, p = .232, \eta_p^2 = .020$. Percentage error was lower for Numbers 7 (2.4%, $SE = 0.45$) and 8 (2.8%, $SE = 0.47$) compared to Number 9 (5.5%, $SE = 0.70$), indicating that there was an end effect. There was a significant main effect of Side on percentage error. Percentage error was lower for Left hand responses (3.0%, $SE = 0.30$) as compared to Right hand responses (3.7%, $SE = 0.32$). The interaction of Number and Side was significant (Fig 5.10). The percentage errors for Number 9 for left hand responses were significantly greater than all of the other responses.

Table 5.9. Experiment 3: Source of repeated measures ANOVA for percentage of errors.

Source	<i>df1</i>	<i>df2</i>	<i>F</i>	<i>p</i>	η_p^2
Arithmetic skill (low, high)	1	69	2.59	.112	.036
Number (1 through 9)	6.9	472.7	2.83**	.007	.039
Number x Skill	6.9	472.7	0.65	.709	.009
Side (left, right)	1	69	5.94*	.017	.079
Side x Skill	1	69	0.03	.865	<.001
Number x Side	6.0	413.0	5.30***	<.001	.071
Number x Side x Skill	6.0	413.0	1.05	.395	.015

* $p < .05$, ** $p < .01$, *** $p < .001$

There is a significant difference between percent errors for left and right hand responses for Number 1, with fewer errors for left- than for right-hand responses (Fig 5.7). This crossover pattern for the end points is consistent with a categorical SNARC effect (Woods et al., 2008). As expected, however, there was no interaction with skill.

Figure 5.7. Experiment 3: Interaction plot of Numbers by Side of response (Yes = right, Yes = left) for Parity on mean percent errors. The error bars are 95% CIs.

ANOVA on Mean Response Times.

A 2(Gender: Male, Female) by 2(Starting button: Left, Right) by 2(Location: on campus, off campus) Univariate ANOVA was conducted with Parity Judgment Response Time as the dependent variable. There were no significant effects and so gender, starting button, and location of experiment were collapsed as factors in the remaining analysis.

Correct responses were analyzed in a 9(Number: 1, 2, 3, 4, 5, 6, 7, 8, 9) by 2(Side of response: Left, Right) by 2(Arithmetic Skill: low, high) mixed ANOVA. Arithmetic skill was a between-subjects factor. A Greenhouse-Geisser epsilon adjustment of degrees of freedom was used if the assumption of homogeneity of variance was violated. The Sidak adjustment was used for multiple pair-wise comparisons.

Table 5.10. Experiment 3: Source for the within-subject ANOVA on mean response times for the Parity Judgment Task.

Factor	<i>df1</i>	<i>df2</i>	<i>F</i>	<i>p</i>	η^2
Arithmetic Skill (low, high)	1	9	<0.01	.965	<.001
Number (1-9)	6.6	453.8	3.50	.001	.048
Number x Skill	6.6	453.8	2.12	.044	.030
Side (left, right)	1	69	0.89	.348	.013
Side x Skill	1	69	1.03	.314	.015
Number x Side	4.5	310.1	4.89	<.001	.066
Number by Side x Skill	4.5	310.1	0.83	.518	.012

** $p < .01$, *** $p < .001$

Figure 5.8. Experiment 3: Mean response times by Number for Parity Judgments. Error bars are 95% CIs.

As shown in Table 5.10, latencies varied across Number (Fig 5.8). Responses to the number 9 (649 ms) were slower than to number 4 (613 ms, $p = .006$) and number 8

(613 ms, $p = .030$). There was an interaction between Number and Arithmetic skill as shown in Figure 5.9. The high-skill group took longer to respond to the numbers 2 and 9 than the low-skill group.

Figure 5.9. Experiment 3: Interaction plot for mean response times of Number by Arithmetic skill for Parity Judgments. Error bars are 95% CIs.

The expected interaction between Number and Side of Response was significant: Participants responded more quickly with the left hand than the right hand on number 1 but responded more quickly with the right than the left hand on number 8 (Fig. 5.10).

Figure 5.10. Experiment 3: Interaction plot of Number by Side of Response for mean Response Times on Parity Judgments. Error bars are 95% CIs.

SNARC effect for Parity.

Regressions were conducted on the difference in left and right hand response times across the number sets for each participant to test for a SNARC effect. As the SNARC effect for parity is known to produce a negative slope for individuals who read from left to right, a one-tailed single-sample t -test was used to evaluate whether the mean slopes were different from zero. The single-sample t -test of the individual slopes was significant for the overall group, as well as for both arithmetic skill groups (Table 5.11). The percentage of participants whose slopes were negative was close to 70% for all three comparisons. An independent t -test was conducted on the slopes for the low and high arithmetic groups. There was no significant difference between the slopes for the two arithmetic skill groups, $t(69) = 1.23, p = .222$.

A paired-samples t -test was conducted between to compare participants' response latencies on the Sequence and Parity judgment tasks. Sequence judgment (945 ms, $SE = 28.7$) took significantly more time than Parity judgment (624 ms, $SE = 11.0$), $t(70) = 13.68$, $p < .001$.

Table 5.11. Experiment 3: Summary of single-sample t -tests on the slopes for the Parity Judgment Task.

Group	M (Beta)	% neg ^a	t	df	p
All Participants	-10.0	69.0	-3.33***	70	<.001
Low Skill	-6.27	68.6	-1.44*	34	.040
High Skill	-13.67	69.4	-3.73***	35	<.001

^a Percentage of slopes that was negative across individuals.

* $p < .05$, *** $p < .001$ (one-tailed)

Odd and Even.

For the parity task, Dehaene et al. (1993) graphed the SNARC effect such that separate lines of best fit were calculated for the odd and the even numbers. Following up on the zigzag pattern of Figure 5.8, response time differences between odd and even numbers (parity) were analysed. A Univariate analysis was conducted on mean response times for odd and even numbers. Response times for Odd numbers (630 ms, $SE = 5.6$) and even numbers (617 ms, $SE = 6.2$) were not significantly different, $F(1, 637) = 2.47$, $p = .117$, $\eta^2 = .004$. A paired-samples t -test was conducted on the individual slopes for odd and even numbers. The mean slope for odd (-7.9, $SE = 2.3$) and even (-8.4, $SE = 3.0$) were not significantly different, $t(70) = 0.17$, $p = .868$. Dehaene et al. (1993) never tested whether the separate slopes for odd and even numbers were the same or different. They

did plot the mean differences separately for the odd and even numbers. For historical purposes lines of best fit were used for odd and even (Fig. 5.11).

Figure 5.11. Experiment 3: Scatterplot of the difference in Response Times for the Parity task, for the Odd numbers (black circles) and the Even numbers (orange diamonds). Lines of best fit are shown.

Analysis for Parity task

A post hoc power analysis, using MorePower 6.0.1 (Campbell & Thompson, 2012) was conducted on the SNARC effects (Appendix A, Table A.6). Power for the parity task with all of the participants was .91. Power for the low arithmetic skill group was .28. Power for the high skill group was .89.

Discussion for Parity.

A SNARC effect was found for the Parity Task for both arithmetic skill groups and the slopes did not differ across skill groups. Thus, consistent with Cipora et al. (2013), but contrary to Hoffman et al. (2014), the SNARC effect within the Parity task was not related to participants' arithmetic ability (as measured by the Calculation Fluency

Test). The same pattern of differences in response times as found Dehaene et al. (1993) was found in the present parity task. A single line of best fit is now used for the mean differences of the full set of stimuli.

Correlations between calculation fluency, and slopes for Parity and Sequence

Because SNARC effects were found for Parity and the Sequence Judgment tasks correlations were calculated between the calculation scores and the individual slopes for the two tasks (Table 5.12). Correlations between slopes for these tasks and arithmetic skill were also included, to evaluate patterns more closely.

Table 5.12. Experiment 3: Correlations among Calculation fluency scores and slopes for the Parity and Sequence tasks ($n = 71$).

Slopes	Calculation fluency	Parity	Ordered sequences
1. Calculation fluency scores	-		
2. Parity	-.157	-	
3. Ordered sequences	-.255*	.230+	-
4. Unordered sequences	-.067	-.255*	.104

+ $p = .054$, * $p < .05$

There was a significant negative correlation between Calculation Fluency score and the slopes of the ordered sequences. As Arithmetic skill increased the slope of the SNARC effect for ordered sequences became more negative (Fig. 5.12). This pattern is consistent with the findings for the SNARC effect in ordered sequences when skill groups were analyzed (Table 5.6).

Figure 5.12. Experiment 3: Scatterplot of slopes for ordered Sequences by Arithmetic skill. Line is line of best fit ($n = 71$).

In contrast, neither the slopes in the parity task nor the slopes for unordered sequences were significantly related to fluency scores. There was, however, a significant negative correlation between slopes of the unordered sequences and the slopes of the parity task. The correlation between parity slopes and ordered sequence slopes approached significance ($p = .054$). Thus, negative slopes in the parity task, which have been the most common way of assessing the SNARC effect (Woods et al., 2008), were related to negative slopes for ordered sequences, but to positive slopes for unordered sequences (Figure 5.13).

Figure 5.13. Experiment 3: Scatterplot of slopes for unordered Sequences by slopes for Parity task. Line is line of best fit ($n = 71$).

The positive correlation between the slopes of the ordered sequences and the parity task suggests the possibility that the slopes in the two tasks (ordered sequence and parity) could be related to the same underlying cognitive process. The scatterplot for the relation between these two slopes is shown in Figure 5.14. As slope for ordered sequence increased, so too did the slope for parity. However, as shown in Figure 5.14, there was one individual with very large slope values on both parity and ordered sequence.

Figure 5.14. Experiment 3: Scatterplot of slopes for ordered Sequences by slopes for Parity task showing best fitting line ($n = 71$).

This outlier may have been influencing the relation between the two variables (ordered sequence and parity). Accordingly, this outlier was removed and the correlations between parity and ordered sequence and between parity and unordered sequence were both recalculated (Table 5.13). Removing the single outlier had a rather dramatic effect on the pattern of correlations. First, arithmetic skill was no longer significantly correlated with slopes on the ordered sequences, although the trend remained. Second, the correlation between slopes of the parity and unordered sequence judgments remained significant (Fig. 5.16). However, the slopes of the parity and ordered sequence judgments were no longer correlated. This one participant seems to have exerted substantial influence on the pattern of data. These extremely variable results suggest that the slope values are not particularly stable indices of participants' performance.

Table 5.13. Experiment 3: Correlation table between Calculation fluency scores and individual slopes for the Parity and Sequence tasks ($n = 70$, outlier removed).

Slopes	Calculation fluency	Parity	Ordered Sequences
1. Calculation fluency scores	-		
2. Parity	-.102	-	
3. Ordered sequences	-.206+	.057	-
4. Unordered sequences	-.050	-.328**	.063

+ $p = .087$, ** $p < .01$,

Figure 5.15. Experiment 3: Scatterplot of slopes for unordered Sequences by slopes for Parity task. Line is line of best fit.

Summary of Correlations.

An initial analysis found a correlation between ordered sequences and arithmetic skill. As arithmetic skill increased the slope progressively changed from positive to negative. A correlation was found between ordered sequences and parity. As the slope of

parity increased so too did the slope of ordered sequences. And a negative correlation was found between the slopes for parity and unordered sequences. As the slope of parity increased the slope of the unordered sequences decreased.

However, the negative correlation between calculation score and ordered sequence slopes shown in Figure 5.12 may have been driven by the data of one individual, whose slopes were very high in both the sequence judgment and parity tasks but whose arithmetic skill was quite low (see Figure 5.12 – outlier is in the top left corner). With the data from the one participant removed, there was a significant negative correlation between the slopes of the unordered sequence task and the slopes of the parity task. As slopes for parity task increase, slopes for unordered sequences decreased (Fig. 5.15).

In general, the slopes for the low scorers seem to be scattered evenly above and below zero. This has made the correlations susceptible to change by a small number of people. Hence, the correlations may be illusory.

Discussion

A SNARC effect was found for the ordered Triads in the Sequence Task for the high arithmetic skill group. This finding replicates Experiment 2 and provides support for the hypothesis that order and not magnitude underlies spatial-numerical associations. However, the correlation between calculation fluency score and the regression slopes for ordered sequences was only marginally significant. Thus, the SNARC effect for ordered sequences was still not stable, despite the relatively large sample size. One possibility was that there were relatively few participants with “high” levels of arithmetic skill. This issue will be addressed in the General Discussion.

A stable SNARC effect was found for Parity, indicating typical performance for the participants in the task that traditionally shows spatial-numerical associations.

Arithmetic skill did not interact with the SNARC effect for Parity, indicating that it was equally present for low and high arithmetic skill groups. The presence of the SNARC effect for the low-skill group in the Parity Task demonstrates that spatial numerical associations are being activated.

There was no spatial numerical association for the unordered stimuli, even for stimuli that participants had ordered during the training task. Nor was there a correlation between calculation fluency scores and the regression slopes for unordered sequences. Hoffman et al. (2014) stated that there were three factors that differentiated low and high arithmetic skill groups in regards to spatial numerical associations. First, the low-skill group would have less automatic access to number facts than the high-skill group. Secondly, the low-skill group may not have used direct recall of parity, instead using a procedure that may have incurred higher working memory loads. Lastly, higher load may have interfered with the ability to inhibit irrelevant number information. In the Sequence Judgment Task, participants had to hold the instructions in working memory, encode three numbers, judge whether they were in order, and press the correct number. The Parity Judgment Task only required the encoding of a single number. Accordingly, the Sequence Judgment Task was more difficult than the Parity Judgment Task, as demonstrated by the significantly longer latencies for the Sequence Judgment Task. The lack of an SNARC effect for ordered sequences in the low Math skill group may therefore reflect differences in task difficulty.

Limitations of Experiment 3

One limitation of Experiment 3 was that the numbers of ordered and unordered stimuli were unbalanced. Equating the number of ordered, unordered-1, and unordered-2 Triads, resulted in twice as many unordered as ordered stimulus sets. This may have created a bias for participants to respond “no”, and a corresponding slowing down of “yes” responses. Evidence for such an effect can be found by comparing latencies across experiments. The difference between ordered and unordered stimuli in Experiment 2 was 90 ms as compared to 45 ms in Experiment 3. The effect size for the stimulus type effect was also reduced, from .303 to .167. A greater percentage of unordered stimuli, compared to ordered stimuli, may have increased reliance on serial comparisons (Bourassa, 2014) and reduced reliance on direct access. This issue will be further addressed in the General Discussion.

The lack of a SNARC effect for unordered sequences may have been due to confusability among the sequences that increased variability, difficulty of task, or the use of processes other than memory recall. Or it could be that the training task was not sufficient to induce order amongst the unordered triads. The number of times the cards are to be sorted could be increased. As well the training could use other tasks, such as asking which triad comes before (or after) the displayed triad.

There were only 16 control participants. When they were further divided into low and high arithmetic groups, the number became even smaller. For future experiments, there should be more participants for each cell of the design. In Experiment 2, the power calculation reported 14 per cell were needed for a power of 80%.

CHAPTER 6 – GENERAL DISCUSSION

Many researchers have observed relations between number and space (reviewed by Wood et al., 2008). These spatial-numerical associations (SNAs) support a model of mental representations of number in which smaller numbers are associated with the left side of space whereas larger numbers are associated with the right side of space. In this thesis, I used a sequence judgment task (e.g., Lyons & Beilock, 2009) as a novel platform for exploring the source of the spatial-numerical association of response codes or SNARC. The SNARC effect (Dehaene et al., 1993) is a specific case in which spatial relations influence performance in numerical tasks that are structured as two-alternative, forced-choice decisions. The overall goal of this thesis was to use the sequential judgment task to determine the relative contributions of order versus magnitude as the source of the SNARC effect. Number representations include both order and magnitude information (Sury & Rubinsten, 2012), and so either or both could contribute to SNAs in numerical tasks. By using a novel numerical task to assess SNARC, I hoped to disentangle ordinal representations from magnitude representations and thus further illuminate the source of SNAs in numerical tasks.

Dehaene et al. (1993) stated that the SNARC effect was not caused by an “overlearned sequential structure” (p. 381). Instead, they concluded that magnitude was automatically accessed, even though it was irrelevant to the task, and was therefore the source of the spatial-numerical effect. Contrary to this conclusion, Gevers et al. (2003, 2004) found the SNARC effect for well-known non-numeric stimuli, specifically, months, days of the week, and letters of the alphabet. Although it could be argued that the months and days are also numerically coded (i.e., January is the first month, December is

the 12th month), letters are not. Hence, Gevers et al.'s results suggested that for non-numeric stimuli, order was being automatically accessed. Nevertheless, the source of the SNARC effect remains somewhat controversial and thus there are different theories about its source, including the reading instruction account (Dehaene et al., 1993; Göbel et al., 2011; Zebian, 2005), the strategic view (Fischer, 2006; van Dijck & Fias, 2011), the brain lateralization account (de Hevia et al., 2014), and the implicit instruction account (Patro et al., 2015). Furthermore, some researchers (e.g., Dodd et al., 2008) have proposed that the SNARC effect in numerical tasks is special, because numbers inherently carry both magnitude and ordinal information. By choosing to explore SNARC effects in the order judgment task, I was able to extend our understanding of SNAs beyond their effects in parity and magnitude comparisons tasks.

Summary of Findings

The first question I addressed in this research was whether SNARC effects would occur in the order judgment task. In Experiment 1, the SNARC effect was found for ordered numeric stimuli for both ascending and descending sequences. There were no SNARC effects for unordered sequences. Critically, however, the effect was most stable and consistent for individuals whose spoken language was Chinese, Japanese, or Korean (and hence, they had been educated in one of these countries). Individuals whose first language was English showed a typical SNARC effect only on descending sequences, and a reversed SNARC on ascending sequences. Participants whose first language was Arabic did not show a significant SNARC effect, although the trend was for positive slopes in both ascending and descending conditions. As there were only seven participants in this language group, power was low. In summary, there was some

evidence for SNARC effects, however, only with the ordered sequences and mainly for the group with Asian first languages. None of the effects for unordered sequences were significant. I hypothesized that the differences among “language” groups was probably not an effect of language, *per se*. Both English- and Asian-language groups know languages with left-to-right reading and writing directions. A more likely source of group differences is in other sources of variability across individuals, specifically, arithmetical skill as a potential predictor of performance on the order judgment task (Lyons & Beilock, 2011) and potentially as an influence on SNARC effects (Cipora et al., 2015).

Experiment 2 further addressed the issues of individual differences and whether SNARC effects occur for the unordered sequences. To examine whether the difference in the SNARC effect between the English and Asian first-language groups was due to better knowledge of arithmetic, participants also completed a measure of arithmetic skill, the Calculation Fluency Test, and arithmetic skill was thus used as a variable in further analyses. In terms of stimuli, in Experiment 1 the distance between digits in the sequences was different across the number magnitudes and contributed to item variability. Thus, in Experiment 2, the stimuli were regularized with a distance of one between the three digits of the ordered stimuli. The results of Experiment 2 were similar to those of Experiment 1 in that SNARC effects were found only for ordered stimuli, and only for more-skilled participants. However, in contrast to Experiment 1, a SNARC effect was found only for ascending (ordered) sequences not for descending sequences.

Experiment 3 explored two questions. First, I hypothesized that a SNARC effect would be found for unordered sequences if the participant was trained to “order” the unordered stimulus sets relative to each other. Second, I included a parity task to

determine whether (a) typical SNARC effects would be found for the population and apparatus being used, and (b) on the assumption that, if the Parity task and the Sequence task tapped into the same mental representations, then there should be a correlation between the two sets of regression slopes for the SNARC effect.

In Experiment 3 there was a SNARC effect for the high arithmetic skill group for the ordered stimuli, replicating the patterns found in Experiments 1 and 2. However, there were no significant SNARC effects for the unordered sequences, even when participants were trained to order the specific unordered triads. Furthermore, in contrast to the relatively weak SNARC effects found for the sequence judgment task, a significant SNARC effect was found for the Parity task. However, arithmetic skill was not significantly related to SNARC effects in the parity task. This pattern suggests that SNARC effects in the two tasks may reflect different aspects of participants' numerical representations and/or that the cognitive processes involved in the two tasks are different.

In summary, the results of three studies provide some evidence for a spatial-numerical association that is activated in the sequence judgment task. However, the effect was relatively modest, was found only on ordered sequences, and only for individuals with relatively high arithmetic competence. Thus, the SNARC effect was surprisingly elusive.

Discussion of Findings

Why was the SNARC effect relatively weak in the context of the sequence judgment task? Comparison of the sequence judgment task to other tasks that have shown stable SNARC effects suggests a number of potentially important differences. Wood et al. (2008) recorded that the majority of the reported SNARC experiments have used

parity, magnitude classification (a single number compared against a fixed standard), or magnitude comparisons (pairs of numbers compared against each other). These numerical tasks vary in how many numbers are involved in the decisions, as well as in the nature of the required decisions. Parity requires the judging of a single number. The two magnitude tasks require judging of two numbers. The sequence task requires the judging of three numbers. These differences across tasks may contribute to variability that may obscure SNAs. Consistent with this view, order judgments are much slower than decisions in other tasks. In Experiment 3, for example, participants took much longer to perform the Sequence judgment task (945 ms) than the Parity judgment task (624 ms). Similarly, order judgments are slow relative to pooled mean response times for the meta-analysis (Wood et al.): for parity (536 ms), magnitude comparison (604 ms) and magnitude classification (632 ms). Thus, other factors influencing task performance may have obscured the relatively sensitive SNARC patterns.

The argument that slowed processing time explained the weak SNARC effects in the order task, however, is inconsistent with the more general finding in the literature in which the size of the SNARC effect *increases* with overall response latencies (Wood et al., 2008). Based on this argument, in Experiments 2 and 3, the low arithmetic skill group has slower response times and thus should have been expected to have steeper regression slopes. Instead, there was no SNARC effect, a flat slope, for the low skill groups. Referring back to the meta-analysis, although latencies were larger for the magnitude comparison task as compared to the parity task, the pooled size of the SNARC effect in magnitude judgment tasks was smaller than for the parity task. The strength of the SNARC effect is variable across tasks and within task types (Table 1. Wood et al., 2008).

It also seems that parity tasks provide the most likely context within which to observe stable SNARC effects, a result supported by the healthy SNARC effect found in the parity task in Experiment 3 of this thesis.

Fischer (2006) suggests that spatial associations may not always be activated even in numerical tasks. Rather, people might make strategic decisions due to “recent and current task demands” (p. 1067). Essentially, this view suggests that activation of spatial associations is not automatic. Spatial-numerical associations may vanish or appear as the task demands increase or decrease. Task complexity may increase the aforementioned task demands. Within Experiment 3, sequence judgments took more time than parity judgments. Compared to parity judgment performance, the low-skilled group made more errors (+2.1%) and took more time (+173 ms) than the high-skilled group to respond to sequence judgments. Sequence judgments vary in difficulty and may require a more complex decision process than parity judgments. For example, ordered sequences are familiar counting sequences, and may be recalled directly from memory. The large response time advantage for the ordered triad {1 2 3} indicates that this triad may have a privileged position in memory. Unordered sequences, on the other hand, do not benefit from familiarity. According to Fischer, some of the stimuli would activate spatial associations, and others would not, depending on task demands (whether the trial was an easy judgment or a difficult judgment).

The original hypothesis in the present research was that SNARC effects would only be found for ordered sequences. Consistent with this prediction, SNARC effects were not found in any of the three experiments for unordered sequences. In Experiments

2 and 3, the SNARC effects for the sequence judgment tasks were further limited to those who had scored in the upper half of the Calculation Fluency Test.

Arithmetic Skill

There has been interest in the relationship between parity judgments and arithmetic skill since Dehaene et al. (1993). In the present thesis, interest in arithmetic skill was sparked by the difference in spatial-numerical associations between English and Asian first-language groups. Although Asian languages traditionally used a top-to-bottom writing direction, in the last few decades the Chinese and the Japanese have adopted a left-to-right direction of reading and writing. Thus, language direction could not account for the differences across groups. These groups have been found to differ substantially in arithmetic skill (Campbell & Xue, 2001). In previous research, performance on the sequence judgment task was strongly correlated with arithmetic fluency, at $r = .70$ in Lyons and Beilock (2011) and $r = .60$ in LeFevre and Bourassa (2016; see also LeFevre & Bisanz, 1986). These correlations suggest that access to some aspects of numerical representations that are important to both arithmetic and to sequence judgments occurs in both of these tasks.

In contrast to the strong relations between sequence judgments and arithmetic skill, researchers have not found consistent relations between SNARC effects in parity judgments and skill (cf., Cipora et al., 2013; Hoffman et al., 2014). Similarly, in the present research, performance on the parity task was not related to arithmetic skill. In other research on the SNARC effect, either there has been no relation between mathematical knowledge and SNARC (e.g., Cipora et al., 2013) or occasionally, less-skilled participants have shown more negative (i.e., larger) slopes (e.g., Hoffman et al.,

2014). Thus, parity judgments and sequence judgments may not necessarily tap into the same underlying mental representations.

One concern for the present research, however, is the definition of the skill levels across the present experiments. According to LeFevre, Penner-Wilger, Pyke, Shanahan, and Deslauriers (2014) between 1993 and 2010, Canadian undergraduate arithmetic fluency scores declined by 37% (note that fluency was indexed scores on the addition and subtraction-multiplication subtests of the Kit of Factor Referenced Cognitive Tests; French et al. 1963). In the present Experiment 3, where a similar measure of arithmetic fluency was used (i.e., the CFT; Sowinski et al., 2015), the fluency score for the high-skilled group was significantly lower than the reported mean CFT score for Chinese-educated individuals ($n = 58$) from an undergraduate sample, $t(69) = 3.67, p < .001$. Using the Chinese mean score as a reference, the high arithmetic skill group in the present experiment may not be high skilled after all. The scatter-plot showing the relation between math score and ordered sequence slope (Fig. 5.12), shows that the majority of participants' math scores were below 45, and the slopes were clustered around zero. It is possible that a truly high-skilled group would show a more stable SNARC effect.

Numerosity Code versus Mental Number Line

Under the numerosity code model (Zorzi et al., 2005), the argument that magnitude is automatically accessed when people process numeric information should have resulted in a SNARC effect for both ordered and unordered sequences. According to the numerosity code model “magnitudes are represented as the quantity of units” (Zhao et al., 2012, p. 1) and thus is separate from the concept of order. In the number line model, in contrast, magnitude and order are coded within a single representation and thus may be

inseparable. There is support for the numerosity code model from physiological studies that have shown that magnitude and order activate different parts of the brain (Badets et al., 2015; Zhao et al., 2012) when they are trained independently. In accord with this model, the ordered and unordered number sets {1 2 3} and {2 3 1} share the same magnitude and thus should have activated magnitude to the same extent. Thus, the prediction based on the numerosity code model is that a SNARC effect would be found for both ordered and unordered stimuli. However, in the present research, the SNARC effect was found only for the ordered sequences, but not for unordered sequences. Even with training to induce relative order within the unordered sets there was no evidence of a stable SNARC effect for the unordered sequences. The present results, therefore, do not support the numerosity code model.

Under the number line hypothesis, “numbers are represented as an ordered sequence of input nodes on an oriented analogical number line” (Zhao, et al., 2005, p. 1). Hence, numbers automatically activate a mental representation that includes both magnitude and order. This activated spatial information interacts with decisional and attentional processes required in specific tasks. Within this hypothesis, the spatial positioning of the number set {1 2 3} should be on the leftmost side of the mental number line, and {7 8 9} should be on the rightmost side of the mental number line. The positioning of unordered sequences is not explicitly predicted. Unordered sequences may not activate specific positions on the mental number line, especially, for counting sequences, because they may result in interference between the familiar and the presented patterns in the order judgment task. Nevertheless, to the extent that participants do activate the whole unordered sequence (cf. Bourassa, 2014), the mental number line

model predicts that the SNARC effect should arise out of judgments about the triads of numbers if the activation of these number sets activates the mental number line, either automatically or intentionally.

A Proposed Model

A four-layer model of judgment processing is suggested (Fig. 6.1) to explain the spatial-numerical associations due to differing complexity levels between tasks. The bottom layer of the model is long-term memory. Task is the second level. Procedures are a possible, but not mandatory, third level. The fourth level is response selection. At the level of task, parity judgments require that participants encode a single number. This stimulus activates number knowledge and spatial information related to the number in long-term memory. If the appropriate facts are available (e.g., direct recall of odd or even information for that number), response selection is the next step. However, as well as activating number knowledge, spatial associations are also activated. This spatial association then influences the response selection, causing the rightward or leftward bias that is the SNARC effect. Thus, the SNARC effect is a consequence of spatial information influencing the response stage.

If information about parity is not activated from long-term memory, then the participant has to use a procedural approach to generate a parity code. For example, they may skip count 2, 4, 6 to determine that 6 is an even number, or divide the number by two (e.g., 7 is not divisible by 2, therefore it is odd), to activate a parity code. This process is presumably slower than direct retrieval of parity, which might result in either more interference from irrelevant spatial information (e.g., with skip counting) or less interference (e.g., if the time course of processing is slowed). Such variability may, in

part, help to explain why the relation between arithmetic skill and the SNARC effect in parity tasks is inconsistent. Further research to explore the relation between solution strategies on the parity task and arithmetic skill might be useful in understanding the role of skill in the SNARC effect.

The sequence judgment task requires that three numbers be activated rather than a single number. This difference in encoding may account, in part, for the large overall difference in solution times for parity and sequence judgments. Each number activates its related number knowledge and spatial information. The three numbers in the order judgment task therefore result in more associations being activated directly than in the parity task. Additionally, within the sequence judgment task, any relationships among the three numbers are also activated. If the task-related information needed to make the judgment is available in long-term memory, as in ordered counting sequences, then a suggested response can be submitted directly to the response selection stage. As for the parity task, spatial associations will also be activated, and potentially interfere with response selection. Otherwise, if no information about order is available from long-term memory, a procedure to determine order must be activated determine the final response. Procedures to determine order may include counting (e.g., for 2 3 1 the next number should be 4 instead of 1 according to a counting procedure) or participants may use sequential comparisons (Bourassa, 2014; Turconi et al., 2006). For the sequence 2 3 1, the first comparison would indicate “yes, ordered” but the second comparison not ascending, so the final response “no, unordered” would be sent to the response selection stage. At this stage, spatial information activated from long-term memory could bias the response, as for parity judgments. The main difference seems to be that the order

judgment processes will take longer, especially if procedural strategies are slow and used relatively frequently.

So how does this model predict that SNARC effects will be larger for individuals with higher levels of arithmetic skill with sequence judgments? People with good calculation skills have more number knowledge, and may have richer spatial-numerical associations than people with poor calculation skills. If people with poor arithmetic skills also lack number knowledge (e.g., lack of familiarity with sequences) they may not be able to activate number – space associations, and hence they would be less likely to show SNARC effects. The solid lines indicate the most likely path for each task. Parity judgments are fast, presumably because parity information is likely to be recalled directly from long-term memory. In contrast, sequence judgments may be more likely to require procedures, especially when the task involves only ascending sequences as ordered (Bourassa, 2014) and thus take longer than parity judgments. Information about spatial relations may therefore have more time to accumulate and influence response selection processes.

Figure 6.1. Model of judgment process.

Limitations

The present research found the SNARC effect to be elusive in that it was found only in slope analyses, not in the full ANOVA design, and was only found for individuals with relatively high levels of arithmetic skill. This is in contrast to the stable SNARC effects observed in parity tasks, both in the literature and in Experiment 3 of the present research. One explanation for why spatial-numerical associations were not found consistently in the present research is the nature of the order judgment task. The order judgment task is a more complex task than the parity task, as evidenced by the slower response times and higher error rates. Another explanation for the relative dearth of

SNAs is that unordered sequences cause some level of confusion because although they activate stored information, it has to be discounted in the order judgment process.

Evidence for such confusion is found in the large differences between ordered and unordered sequences for counting (sequential) numbers, but not for sequences such as 2 5 9 which have no prior associations (Bourassa, 2014; Lyons & Beilock, 2009).

Accordingly, participants need to suppress the information in unordered sequences that are derived from familiar sequences, potentially inhibiting number line that might contribute to a SNARC effect. Consistent with this view, no spatial-numerical association was found for the unordered sequences, even after a brief training task.

Another limitation of this research is that the training task was relatively brief and so may not have influenced participants' solution strategies. Participants were asked to locate the triad on a number line, and to sort cards only twice that had the unordered triads printed on them. Though all the participants were able to locate the triads and sort them without any errors, this amount of practice may not have been sufficient to induce a strong relative order that would be activated automatically. In contrast, Previtali et al. (2010) found a SNARC effect after having participants learn sequences of unrelated words. During the learning phase, the ordering of the words was an explicit goal. During the training phase, the order of the number words was tested with "what comes before" and "what comes next" questions. All of the participants experienced the same number of learning and training trials. At the end of the training the mean error rate was 11.6%. That level of training presumably induced relatively strong ordered associations among the stimuli.

Another limitation of this research was that the language effect in Experiment 1 was not definitely linked to skill differences because there were few Asian participants in subsequent experiments. In Experiment 1, a SNARC effect was found for Asian first-language participants. On the basis that arithmetic skill, and not language per se was the differentiating factor, the next two experiments used the Calculation Fluency Test. Because only a small number of Asian first-language participants took part in Experiments 2 and 3, it was not possible to directly test the assertion that arithmetic skill was the differentiating factor in Experiment 1. Future research that included Asian-educated individuals but controlled for arithmetic skill is necessary to determine if that factor accounted for the differences across groups.

Another factor that may have interfered with finding stable SNARC effects was that the composition of stimuli varied across the three experiments. In Experiment 3, in particular, participants judged twice as many unordered stimuli as ordered stimuli. If, as Fischer (2006) suggested, participants are strategic in their task performance, the 2:1 ratio may have created a bias among the participants towards the use of serial comparison procedures rather than direct recall when judging ordered sequences. Support for this effect is found in that the difference between ordered and unordered sequences in Experiment 3 was smaller in size than in the previous experiments. If Fischer is correct that number knowledge and number-space associations are not always automatically activated, stimulus composition could be one source of differential activation of specific information. Thus, the selection of stimuli in Experiment 3 may have reduced the influence of those features that may contribute to SNARC effects in the order judgment task.

Future Research

The parity task involves a single number. A fixation point preceding the stimulus insures that the participant is focused on where the stimulus will appear. Eye motion to capture the number is not needed, nor is it expected. But when three numbers are used in a stimulus, the eye gaze may shift to each number. Eye tracking data may provide useful information as to the scanning of the three numbers and, for example, whether processing patterns might explain the reverse SNARC effect seen in Experiment 1 for the English first-language group. Further, use of attentional bias through requiring participants to move their eyes to one side of the screen or the other before each trial could be used to induce a spatial-numerical association in the same manner that Patro et al. (2015) used a finger slide across a touch screen on a notebook to induce a temporary spatial-numerical association.

Possible manipulations might include increasing the physical distance between the three numbers of the stimulus set, causing the eyes to traverse the screen. Alternatively, or in combination, participants could be asked to scan the sequences from left-to-right, on one half of the trials, and from right-to-left on the other half of the trials. Such a manipulation could also be introduced more subtly by varying whether both ascending and descending sequences were included and scanning patterns could be compared. Participants would be free to adopt their preferred scanning of the numbers, but scanning of the numbers in a congruent direction to the order of the numbers might increase the probability of detecting a spatial-numerical association.

Another possibility for using eye tracking is to add an attentional task during the SNARC experiment as in other SNARC research (Dodd, Van der Stigchel, Leghari,

Fung, & Kingstone, 2008; Fischer, Castel, Dodd & Pratt, 2003). In an attentional SNARC task, the stimulus appears centrally between two target boxes. The participants are told that the stimuli are not predictive of which box in which the target will appear. After each trial, a cue is presented in one of the two boxes. Small numbers bias gaze towards the left box, and large numbers bias gaze towards the right box. This paradigm might show that participants' attention is directed towards the direction of the sequence immediately after each trial, that is, towards the right after an ascending sequence and towards the left after a descending sequence. If such an effect occurs, it might help to explain why SNARC effects are relatively weak in the order judgment task, despite the strong requirement of the task to activate order information.

Conclusion

In conclusion, the results of this thesis suggest that in the order judgment task, effects of spatial-numerical associations are modest and variable across context and characteristics of the participants. Nevertheless, the findings do suggest that further research, using the lessons learned from the current work, could provide interesting and valuable information about SNARC effects in numerical tasks. I conclude that the issue of whether order or magnitude information drives SNARC effects in typical numerical tasks, such as parity or number comparison, remains an open question.

REFERENCES

- Aiken, L. R., & Williams, E. N. (1973). Response times in adding and multiplying single-digit numbers. *Perceptual and Motor Skills*, 37, 3-13.
- Bächtold, D., Baumüller, M., & Brugger, P. (1998). Stimulus-response compatibility in representational space. *Neuropsychologia*, 36(8), 731-735.
- Badets, A., Boutin, A., & Heuer, H. (2015). Mental representations of magnitude and order: A dissociation by sensorimotor learning. *Acta Psychologica*, 157, 164-175.
- Bailey, D. H., Littlefield, A., & Geary, D. C. (2012). The codevelopment of skill at and preference for use of retrieval-based processes for solving addition problems: Individual and sex differences from first to sixth grades. *Journal of Experimental Child Psychology*, 113, 78-92.
- Benoit, L., Lehalle, H., Molina, M., Tijus, C., & Jouen, F. (2013). Young children's mapping between arrays, number words, and digits. *Cognition*, 129, 95-101.
- Berch, D. B., Foley, E. J., Hill, R. J. & Ryan, R. M. (1999). Extracting Parity and Magnitude from Arabic Numerals: Developmental changes in number processing and mental representation. *Journal of Experimental Child Psychology*, 74, 286-308.
- Bourassa, A. (2014). Numerical Sequence Recognition: Is Familiarity or Ordinality the Primary Factor in Performance? Unpublished Masters Thesis, Carleton University, Ottawa, Canada.
- Bull, R., Clelland, A. A., & Mitchell, T. (2013). Sex Differences in the Spatial Representation of Number. *Journal of Experimental Psychology: General*, 142(1), 181-192.

- Campbell, J. I. D., & Thompson, V. A. T. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods*, 44, 1255-1265. doi: 10.3758/s13428-102-0186-0.
- Campbell, J. I. D., & Xue, Q. (2001). Cognitive Arithmetic Across Cultures. *Journal of Experimental Psychology*, 130(2), 299-315.
- Cipora, K., & Nuerk, H. –C. (2013). Is the SNARC effect related to the level of mathematics? No systematic relationship observed despite more power, more repetitions, and more direct assessment of arithmetic skill. *The Quarterly Journal of Experimental Psychology*, DOI:10.1080/17470218.2013.772215
- Cipora, K., Patro, K., & Nuerk, H. –C. (2015). Are Spatial-Numerical Associations a Cornerstone for Arithmetic Learning? The Lack of Genuine Correlations Suggests No. *Mind, Brain, and Education*, 9(4), 190-206.
- Clark, J. M., & Campbell, J. I. D. (1991). Integrated versus modular theories of number skills and acalculia. *Brain and Cognition*, 17(2), 204-239.
- Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, 122(3), 371-396.
- de Hevia, M.D., Girelli, L., Addabbo, M., & Cassia, V.M. (2014). Human infants' preference for left-to-right oriented increasing numerical sequences. *PloS ONE*, 9, e96412. doi:10.1371/journal.pone.0096412.
- Dodd, M. D., Van der Stigchel, S., Leghari, M. A., Fung, G., & Kingstone, A. (2008). Attentional SNARC: There's something special about numbers (let us count the ways). *Cognition*, 108(3), 810-818.

- Fias, W., Brysbaert, M., Geypens, F., & d'Ydewalle G. (1996). The Importance of Magnitude Information in Numerical Processing: Evidence from the SNARC Effect. *Mathematical Cognition*, 2(1), 95-110.
- Fias, W., & Fischer, M. H. (2005). Spatial representation of numbers. In J. I. D. Campbell (Ed), *Handbook of Mathematical Cognition* (pp. 43 – 54). New York: Psychology Press.
- Fischer, M. H. (2004). Orthographic contributions to perceived word centre. *Brain and Language*, 88, 321-330.
- Fischer, M.H. (2006). The future for SNARC could be stark. . . . *Cortex*, 42, 1066–1068.
- Fischer, M. H., & Brugger, P. (2011). When digits help digits: Spatial-numerical associations point to finger counting as prime example of embodied cognition. *Frontiers in Psychology*, 2, 1-7.
- Fischer, M. H., Castel, A. D., Dodd, M. D., & Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nature Neuroscience*, 6(6), 555-556.
- 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, 333-336.
- Fischer, M. H., & Rottmann, J. (2005). Do negative numbers have a place on the mental number line? *Psychology Science*, 37(1), 22-32.
- Fischer, M. H., & Shaki, S. (2014). Spatial biases in mental arithmetic. *The Quarterly Journal of Experimental Psychology*, 67(8), 1457-1460.
- Fischer, M. H., Shaki, S., & Cruise, A. (2009). It takes just one word to quash a SNARC. *Experimental Psychology*, 56(5), 361-366.

- Fitousi, D., Shaki, S., & Algom, D. (2009). The role of parity, physical size, and magnitude in numerical cognition: The SNARC effect revisited. *Attention, Perception, & Psychophysics*, 71(1), 143-155.
- French, J. W., Ekstrom, R. B., & Price, I. A. (1963). *Kit of reference tests for cognitive factors*. Princeton, NJ: Educational Testing Service.
- Fuson, K. C. (1988). Children's counting and concepts of number. Springer-Verlag, New York.
- Gelman, R., & Gallistel, C. R. (1978). The child's understanding of number. Harvard University Press, Cambridge, Mass. 260.
- Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition*, 87(3), 87-95.
- Gevers, W., Reynvoet, B., & Fias, W. (2004). The mental representation of ordinal sequences is spatially organised: Evidence from days of the week. *Cortex*, 40, 171-172.
- Gevers, W., Verguts, T., Reynvoet, B., Caessens, B., & Fias, W. (2006). Numbers and space: A computational model of the SNARC effect. *Journal of Experimental Psychology: Human Perception and Performance*, 32(1), 32-44.
- Göbel, S. M., Shaki, S., & Fischer, M. H. (2011). *Journal of Cross-Cultural Psychology*. 42(4), 543-565.
- Hoffman, D., Mussolin, C., Martin, R., & Schiltz, C. (2014). The Impact of Mathematical Proficiency on the Number- Space Association. *PloS, One* 9(1), 1-11.
- Hubbard, E. M., Piazza, M., Pinel, P. & Dehaene, S. (2005). Interactions between numbers and space in parietal cortex. *Nature Reviews Neuroscience*, 6: 435-448.

Hume, T. (2011). EstimationLine. Software for the iPad.

<http://hume.ca/ix/estimationline.html>

Hung, Y., Hung, D. L., Tzeng, O. J., & Wu, D. H. (2008). Flexible spatial mapping of different notations of numbers in Chinese readers. *Cognition*, 106(3), 1441-1450.

Ito, Y., & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Memory & Cognition*, 32(4), 662-673.

Knops, A., Zitzmann, S., & McCrink, K. (2013). Examining the presence and determinants of operational momentum in childhood. *Frontiers in Psychology*, 4, 1-14.

LeFevre, J.-A., & Bisanz, J. (1986). A cognitive analysis of number-series problems: sources of individual differences in performance. *Memory & Cognition*, 14(4), 287-98. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/3762382>

LeFevre, J. -A., Liu, J. (1997). The role of experience in numerical skill: Multiplication performance in adults from Canada and China. *Mathematical Cognition*, 3(1), 31-62.

LeFevre, J.-A., Penner-Wilger, M., Pyke, A. A., Shanahan, T., & Deslauriers, W. A. (2014). *Putting two and two together: Declines in arithmetic fluency among young Canadian adults, 1993 to 2005*. Technical Report: Math Cognition Laboratory. Ottawa, ON. Retrieved from <http://www.carleton.ca/ics/research/technical-reports/view-reports/>

Link, T., Nuerk, H.-C., & Moeller, K. (2014). On the relation between the mental number line and arithmetic competencies. *The Quarterly Journal of Experimental Psychology*, 67, 1597-1613.

- Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(1), 149-157.
- Lyons, I. M., & Beilock, S. L. (2009). Beyond quantity: Individual differences in working memory and the ordinal understanding of numerical symbols. *Cognition*, 113, 189-204.
- Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition*, 121, 256-261.
- Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1-6. *Developmental Science*, 1–13. doi:10.1111/desc.12152
- Masson, N., & Pesenti, M. (2014). Attentional bias induced by solving simple and complex addition and subtraction problems. *The Quarterly Journal of Experimental Psychology*, 67(8), 1514–1526.
- Mix, K. S., Sandhofer, C. M., & Baroody, A. J. (2005). Number words and number concepts: The interplay of verbal and nonverbal processes in early quantitative development. In R. V. Kail (Ed.) *Advances in Child Development and Behavior*, 33, 305-346. New York, Elsevier.
- Miller, K., & Gelman, R. (1983). The Child's Representation of Number: A Multidimensional Scaling Analysis. *Child Development*, 54(6), 1470-1479.
- Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature*, 215, 1519-1520.

- Nieder, A., & Miller, E.K. (2004). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 7457–7462.
- Notebaert, W., Gevers, W., Verguts, T., & Fias, W. (2006). Shared Spatial Representations for Numbers and Space: The Reversal of the SNARC and the Simon Effects. *Journal of Experimental Psychology: Human Perception and Performance*, 32(5), 1197-1207.
- Patro, K., Fischer, U., Nuerk, H. -C., & Cress, U. (2015). How to rapidly construct a spatial-numerical representation in preliterate children (at least temporarily). *Developmental Science*, 1-19. doi: 10.1111/desc.12296.
- Prado, J., Van der Henst, J., & Noveck, I. A. (2008). Spatial associations in relational reasoning: Evidence for a SNARC-like effect. *The Quarterly Journal of Experimental Psychology*, 61(8), 1143-1150.
- Previtali, P., de Hevia, M. D., & Girelli, L. (2010). Placing order in space: The SNARC effect in serial learning. *Experimental Brain Research*, 201(3), 599-605.
- Restle, F. (1970). Speed Of Adding And Comparing Numbers. *Journal of Experimental Psychology*, 83(2), 274-278.
- Rubinsten, O., & Sury, D. (2011). Processing ordinality and quantity: The case of developmental dyscalculia. *PloS One*, 6, 1-12.
- Santens, S., & Gevers, W. (2008). The SNARC effect does not imply a mental number line. *Cognition*, 108, 263-270.

- Shaki, S., Fischer, M. H. & Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychonomic Bulletin & Review*, 16 (2), 328-331.
- Shaki, S., & Leth-Steensen, C. (2009). *Instruction-dependent SNARC for comparisons of linearly ordered symbolic stimuli*. In the Proceedings of the 25th Annual Meeting of the International Society for Psychophysics, Galway Bay, Ireland.
- Shaki, S., & Petrusic, W. M. (2005). On the mental representation of negative numbers: Context-dependent SNARC effects with comparative judgments. *Psychonomic Bulletin & Review*, 12(5), 931-937.
- Shaki, S., Petrusic, W. M., Leth-Steensen, C. (2012). SNARC effects with numerical and non-numerical symbolic comparative judgments: Instructional and cultural dependencies. *Journal of Experimental Psychology: Human Perception and Performance*, 38(2), 515-530. doi: 10.1037/a0026729
- Shimizu, H. (2002). Measuring keyboard response delays by comparing keyboard and joystick inputs. *Behavior Research Methods, Instruments and Computers*, 34(2), 250-256.
- Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *The origins of cognitive skills* (pp. 229-293). Hillsdale, NJ: Erlbaum.
- Sowinski, C., Dunbar, K., & LeFevre, J. (2015). *Calculation Fluency Test*. Unpublished technical report, Math Lab, Carleton University, Ottawa, Canada. Retrieved from: <http://carleton.ca/cacr/math-lab/measures/calculation-fluency-test/>

- Sudevan, P., & Taylor, D. A. (1987). The cueing and priming of cognitive operations. *Journal of Experimental Psychology: Human Perception and Performance*, 13(1), 89-103.
- Sury, D. & Rubinsten, O. (2011). Ordinal Processing of Numerical and Non-numerical Information. Z. Breznitz et al. (eds.), *Reading, Writing, Mathematics and the Developing Brain*, 209, 209-232.
- Turconi, E., Campbell, J. I. D., & Seron, X. (2006). Numerical order and quantity processing in number comparison. *Cognition*, 98(3), 273-285.
- Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology*, 23(4), 515-557.
- van Opstal, F., Fias, W., Peigneux, P., & Verguts, T. (2009). The neural representation of extensively trained ordered sequences. *Neuroimage*, 47, 367-375
- van Dijck, J -P, & Fias, W. (2011). A working memory account for spatial numerical associations. *Cognition*, 119, 114-119.
- Wood, G., Nuerk, H. -C, & Willmes, K. (2006). Variability of the SNARC effect: systematic interindividual differences or just random error? *Cortex*, 42, 1119-1123.
- Wood, G., Willmes, K., Nuerk, H.-C., & Fischer, M. H. (2008). On the cognitive link between space and number: A meta-analysis of the SNARC effect. *Psychology Science Quarterly*, 4(4), 489–525. doi:10.1027/1618-3169.52.3.187
- Zebian, S. (2005). Linkages between number concepts, spatial thinking, and directionality of writing: The SNARC effect and the REVERSE SNARC effect in English and Arabic monoliterates, biliterates, and illiterate Arabic speakers.

Journal of Cognition and Culture. Special Issue: On Psychological and Cognitive Foundations of Religiosity, 5(1-2), 165-190.

Zhao, H., Chen, C., Zhang, H., Zhou, X., Mei, L., Chen, C., Chen, L., Cao, Z., & Dong, Q. (2012). Is order the defining feature of magnitude representation? An ERP study on learning numerical magnitude and spatial order of artificial symbols. *PloS One*, 7, 1-9.

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, 674-680.

Zorzi M, Stoianov I, Umiltà, C. (2005) Computational Modeling of Numerical Cognition. In: Campbell J, editor. Handbook of Mathematical Cognition. New York: Psychology Press. pp. 67–84.

APPENDICES

Appendix A: Power Analyses

Table A.1. Summary of power analysis for Experiment 1.

Language	Task	Type	<i>SD</i>	<i>t</i>	<i>n</i>	η^2	power
English							
	ascending	unordered	36.3	-0.66	11	.04	.07
		ordered	21.0	2.75**	11	.43	.69
	descending	unordered	58.3	-1.00	11	.09	.12
		ordered	26.2	-2.77**	11	.43	.70
Asian							
	ascending	unordered	36.0	-0.82	13	.25	.10
		ordered	25.5	-2.92**	13	.42	.76
	descending	unordered	82.6	-1.10	13	.09	.15
		ordered	32.2	-1.73+	13	.20	.33
Arabic							
	ascending	unordered	55.7	0.63	7	.06	.06
		ordered	151.8	0.57	7	.05	.05
	descending	unordered	135.1	0.08	7	<.01	.03
		ordered	99.6	1.09	7	.17	.11

+ $p = .109$, ** $p < .01$ (one-tailed)

Table A.2. Summary of power analysis for Experiment 2.

Skill	Task	<i>SD</i>	<i>t</i>	<i>n</i>	η^2	power
Low	ascending ordered	18.1	-1.07	13	.09	.14
	ascending unordered	22.7	0.39	13	.01	.05
	descending ordered	29.6	0.46	13	.02	.06
	descending unordered	30.2	-0.67	13	.04	.08
High	ascending ordered	18.4	-2.51*	13	.34	.63
	ascending unordered	16.0	-1.84+	13	.22	.37
	descending ordered	21.9	-0.40	13	.01	.05
	descending unordered	32.8	-0.10	13	<.01	.03

+ $p = .063$

Table A.3. Summary of power analysis for ordered sequences Experiment 3.

Group	Stimulus Type	<i>SD</i>	<i>t</i>	<i>n</i>	η^2	power
All participants	ordered	33.2	-1.58	73	.03	.34
Low Skill	ordered	36.6	0.62	37	<.01	.03
High Skill	ordered	28.4	-2.71**	36	.17	.75

** $p < .01$ (two-tailed)

Table A.4. Summary of power analysis for unordered sequences for Experiment 3.

Skill	Training Cards ^a	Stimulus	<i>SD</i>	<i>t</i>	<i>n</i>	η^2	power
Low	Type 1	unordered-1	30.6	-1.76+	19	.15	.37
		unordered-2	34.2	0.79	19	.03	.10
	Type 2	unordered-1	26.8	0.68	10	.05	.07
		unordered-2	23.6	0.98	10	.10	.12
High	Type 1	unordered-1	21.7	0.74	12	.05	.09
		unordered-2	33.9	1.04	12	.09	.14
	Type 2	unordered-1	21.0	1.00	16	.06	.14
		unordered-2	25.7	0.45	16	.01	.06

+ $p < .095$ (two-tailed)

^aType 1 cards matched unordered-1 Triads. Type 2 cards matched unordered-2 Triads.

Table A.5. Summary of power analysis for unordered sequences for controls for Experiment 3.

Skill	Stimulus	<i>SD</i>	<i>t</i>	<i>n</i>	η^2	power
All	unordered-1	45.4	-0.20	16	<.01	.04
All	unordered-2	20.1	-0.32	16	<.01	.05
Low	unordered-1	62.1	-0.23	8	<.01	.03
	unordered-2	24.2	0.07	8	<.01	.03
High	unordered-1	23.1	0.05	8	<.01	.03
	unordered-2	16.4	-0.67	8	.06	.07

Table A.6 Summary of power analysis for parity for Experiment 3.

Group	<i>SD</i>	<i>t</i>	<i>n</i>	η^2	power
All Participants	25.4	-3.33***	71	.14	.91
Low Skill	25.8	-1.44*	35	.06	.28
High Skill	24.8	-3.73***	36	.24	.89

* $p < .05$, *** $p < .001$ (one-tailed)

Appendix B – Extra Experiment 1 figures

Figure B.1. Interaction plot of Response Times for Triad by Side of response for the unordered stimuli for English group. The error bars are CI 95%

Figure B.2. Interaction plot of Response Times for Triad by Side of response for the ordered stimuli for English group. The error bars are CI 95%

Figure B.3. Interaction plot of Response Times for Triad by Side of response for the ordered stimuli for Arabic group. The error bars are CI 95%

Figure B.4. Interaction plot of Response Times for Triad by Side of response for the ordered stimuli for Arabic group. The error bars are CI 95%

Appendix C - Previous SNARC experiments

Table C.1 Summary of experiments using non-numeric stimuli and the factors tested listed by date

Non-numeric	Factors / experiment	Researchers
	Classification task	Previtali, de Hevia & Girelli (2010)
	Relational	Prado, Van der Herst & Noveck (2008)
	Stimulus size, duration	Xuan, Zhang, He & Chen (2007)
	Days of the week	Gevers, Reynvoet & Fias (2004)
	Letters and months	Gevers, Reynvoet & Fias (2003)

Table C.2 Summary of experiments involving synaesthetes and the factors tested listed by date

Synaesthesia	Factors / experiment	Researchers
	Case study	Hubbard et al. (2009)
	Case studies	Jarick et al. (2009)
	Order	Price & Mentzoni (2008)
	Colour	Sagiv et al. (2006)

Table C.3 Summary of experiments using numbers and the factors tested listed by date

Numeric	Factors / experiment	Researchers
	Math skill	Hoffman et al. (2014)
	Visuo-spatial	Viarouge, Hubbard & McCandliss (2014)
	Embodied Cognition	Marghetis, Núñez & Bergen (2014)
	Math skill	Cipora & Nuerk (2013)
	Math skill – Gender	Bull, Cleland & Mitchell (2013)
	Recipes / Language	Fischer, Mills & Shaki (2010)
	Visuo & verbal-spatial	Gevers et al. (2010)
	Dyslexia	Göbel & Snowling (2010)
	Language, spoken vs. read	Fischer, Shaki & Cruise (2009)
	Semantic Congruity Effect	Fischer, Shaki & Algom (2009)
	Response duration not RT	Kiesel & Vierck (2009)
	Switching costs	Liefvooghe & Verbruggen (2009)
	Free response (either side)	Daar & Pratt (2008)
	Second vs. first language	De Brauwer, Duyck & Brysbaert (2008)
	Attentional	Dodd et al. (2008)
	Asian languages	Hung, Hung, Zheng & Wu (2008)
	ANS, number line	Izard & Dehaene (2008)
	Close-far, no number line	Santens & Gevers (2008)
	Reading, Listening	Shaki & Fischer (2008)
	Elementary school	Van Galen & Reitsma (2008)
	Two-digit numbers, matching	Zhou, Chen, Chen & Dong (2008)
	Vertical and horizontal	Müller & Shwarz (2007)
	Stimulus size, duration	Xuan, Zhang, He & Chen (2007)
	Magnitude Comparison and Simon effects	Notebaert, Gevers, Verguts & Fias (2006)
	Size congruity effects	Choplin & Logan (2005)
	Attentional	Fischer, Castel, Dodd & Pratt (2003)
	Elementary school	Berch, Foley, Hill & Ryan (1999)
	Spatial frame of reference	Bächtold et al. (1998)